

Climate come-uppance delayed

A proposal to control greenhouse gases may have been dead on arrival in the US Senate — but its time will surely come.

A modest target for the stabilization and subsequent reduction of greenhouse-gas emissions. Legal authority for the Environmental Protection Agency to regulate such emissions. Establishment of some of the rules needed to foster a free and open market for emissions trading. Some fine-tuning of the federal government's climate-change research programme.

The central points of the American Investments for Reduction of Emissions Act of 2003, which was introduced last week by senators John McCain (Republican, Arizona) and Joe Lieberman (Democrat, Connecticut), would be regarded by many as a perfectly reasonable set of precautions against the likely dangers of global climate change.

Will the US Senate enact the measure? Not a chance. After due consideration at a hearing of the commerce committee on 8 January, the measure is going nowhere fast in the Senate. The Bush administration is dismissive of it and the House of Representatives is uninterested.

Two of Bush's most formidable opponents — McCain inside the Republican party, Lieberman outside it — know this well enough. Their real intention is not to pass a bill. It is to send a signal that the president's nonchalant disregard of this issue will one day come back to haunt him (see page 202).

McCain picked up his own interest in the issue on the campaign trail for the last presidential election. In the cold school halls of Iowa and New Hampshire, the environmentally hawkish senator gained the impression that young people in America feel betrayed by

Washington's failure to engage with the problem of global warming. He changed tack on the issue, and now joins long-time environmentalist Lieberman in addressing it.

But the two men are also sending out a message to the world at large: that Bush doesn't speak for America on climate change. Their concerns are quietly shared by many Americans — in fact by most, according to polls. And around America, two-thirds of states are taking measures that will encourage a reduction in greenhouse-gas emissions.

Meanwhile, elsewhere in the world, ratification of the Kyoto Protocol is proceeding far more vigorously than its supporters anticipated — partly on account of the characteristic petulance with which Bush saw fit to withdraw from it. Instead of using his exit as an excuse to renege on the agreement, as they might well have done, close allies with conservative governments have confirmed their plans to ratify. Last month, indeed, Canada became the hundredth nation to do so. Russia is likely to come on board too, ensuring that participation reaches the threshold to bring the treaty into force without US participation.

In the long run, people who have a sense of the appropriate role of the United States in the world — people such as McCain and Lieberman — will regain influence. Something akin to the bill that the dynamic duo introduced last week will one day pass into law. While waiting for that day, the rest of the world must confront the challenge of global warming on its own. ■

More heat, less light on Lomborg

A Danish committee has picked an appropriate target and misfired.

Not surprisingly, last week's ruling by the Danish Committees on Scientific Dishonesty (DCSD) that Bjørn Lomborg, in his controversial book *The Skeptical Environmentalist*, selected data in a "severely biased" manner and exhibited poor scientific practice (see page 201) received widespread international media coverage. But whether the DCSD emerged with credit also deserves reflection.

Lomborg's hypothesis that warnings issued by environmentalists and scientists are unwarranted, presented in the book rather than in the peer-reviewed literature, has been widely criticized by researchers. But what is the DCSD's authority to tackle what many consider a polemical rather than scientific book?

The DCSD was the first European body to be set up — by the Danish Research Agency — to examine issues of scientific misconduct, and it is still unusual in being mandated to consider any complaint about any scientist, or any scientific work, emerging from both the private and public sectors. A look at its guiding principles (see <http://www.forsk.dk/eng/index.htm>) and its judgement (see www.forsk.dk/uvvu/nyt/udtaldebat/bl_decision.htm) confirms that the DCSD has the freedom to assess the case because, arguably, Lomborg presented himself as an academic and his book as a scientific argument. Appropriately enough, the DCSD emphasizes that it is assessing Lomborg's scientific standards, not his conclusions.

The national context of this independent assessment is relevant here. Lomborg was made director of the politically influential Danish

Environmental Assessment Institute, founded by the new right-wing government after the 2001 elections, solely on the strength of it. According to its own statutes, the institute must be headed by a scientist of appropriate research experience, whereas Lomborg has little additional experience.

Lomborg's claims in his book are certainly significant and potentially influential. The Danish public, at least, has the right to know whether he is arguing on scientifically rigorous grounds, not least given the influence of his position.

Unfortunately, the DCSD has left itself in a weak position. It did not conduct an independent analysis of the book but relied on published criticisms, especially a controversial selection published by *Scientific American*. Even to call this judgement's basis a 'meta-analysis' would be too generous: there is, for example, no justification given for the particular selection of published critiques. Furthermore, through a tangled combination of translation and legalese, the committee's judgement characterizes Lomborg as "objectively dishonest" while at the same time stating that they have no evidence for what most people would call dishonesty: deliberate misrepresentation. That subtle, not to say tortuous, distinction has been lost in the media coverage.

There remains a need for rigorous scrutiny of Lomborg's methods, given his prominence, his claims to serious analysis, and the polarized debate surrounding his book. But this episode leaves everyone little wiser, and the waters surrounding Lomborg even muddier. ■





The Raël thing?
Would-be cloners
hope to publish in
scientific journals
p199



Paris protest
Outcry over call for
Europe to cut science
link with Israel
p200



Sceptical inquiry
Bjørn Lomborg
brought to book
over publication
p201



Skeleton key
Archaeologists
prepare for study of
Kennewick Man
p202

US officials urge biologists to vet publications for bioterror risk

Erika Check, Washington

Biologists need to devise a better process for handling the publication of unclassified research that bioterrorists could use, senior US government officials said last week.

The warning came at a meeting of scientific leaders, national-security professionals and government aides at the National Academies of Science (NAS) on 9 January. Scientific and security groups organized the gathering in response to fears that life-science research could be misused by terrorists.

John Marburger, head of the White House Office of Science and Technology Policy (OSTP), said that the Bush administration believes that most research should be kept unclassified. But he said the administration is searching for ideas about how best to deal with unclassified, 'dual-use' research that seems to aid the cause of terrorists and other would-be makers of bioweapons.

"Society expects its government to take reasonable steps to reduce the risk of bioterrorism, and the question before us is the nature of those steps," Marburger said.

Many of the 200 senior scientists, publishers and officials at the meeting argued anxiously that scientists must be free to publish all unclassified work. But others acknowledged that the community needs to reassure the public and the government that it is acting responsibly — even if they seemed unsure what steps they should take. Editors of journals, including *Nature*, who attended the meeting said afterwards that they hoped to release a joint statement shortly.

Parney Albright, an OSTP official now on secondment to the newly formed Department of Homeland Security, said that biologists need to think harder about the peer-review process. "There's a trade-off between the risks and benefits of efficient and open publication of scientific research, and the community needs to understand that better," Albright said.

He reminded the meeting that public distress over several controversial papers in the past two years has led to calls for restraint or regulation. One example arose in February 2001, when Australian researchers showed that altering a mousepox protein enabled



John Marburger wants to ensure dual-use research on subjects such as smallpox (inset) is kept safe.

it to kill animals that had been vaccinated against the virus. Last June, medical researchers showed how a smallpox protein disables a key element of the human immune system. And last August, Eckard Wimmer and colleagues at the University of New York at Stony Brook published a method for synthesizing poliovirus from scratch.

Albright said these papers had inspired unease among politicians and the public. "The science community ought to come up with a process before the public demands the government do it for them, and that will be driven by the rate at which controversial papers hit the streets," Albright said.

Two of the researchers whose papers were mentioned said they were not sure how this process should work. But neither regretted publishing their studies.

Immunologist Ariella Rosengard of the University of Pennsylvania School of Medicine said that she felt conflicted about publishing her work on smallpox proteins, in the *Proceedings of the National Academy of Sciences* last June. But keeping such results locked up would only slow down the bio-defence effort, she said. "I'm worried that if we eliminate our ability to freely express our research results, we will end up saying it's just not worth it," Rosengard said.

Journal editors said they could think of

few examples of papers that they would absolutely refuse to publish for security reasons. But some have begun including security concerns as a factor in peer review.

For instance, reviewers at the American Society of Microbiology's (ASM) 11 journals can flag a paper that appears to involve a "misuse of microbiology", said Sam Kaplan, chair of the ASM's Publications Board. Editors then review the paper and decide whether to reject it for security reasons.

ASM editors reviewed two papers for security reasons last year, said Kaplan. One is still under review; the other will be published in March. He urged government participants in the meeting not to put new burdens on reviewers. "I hope you all will realize the fragility of this process," Kaplan said.

Some participants pointed out the difficulties of implementing any changes that the United States may wish for in the international world of science. Sam Kaplan noted that 15% of the ASM's editorial board members are from abroad — as are some 60% of papers submitted to the journals.

"Other countries are moving slower on this," said NAS president Bruce Alberts. "We're going to have to have some international effort here or we don't have anything." He plans to raise the matter at a meeting of academy presidents in Amsterdam next week. ■

INSET: SCIENCE/SPL; EUGENE HOSHIKO/AP

Gravity experiment sparks spat between physicists

Geoff Brumfiel, Seattle

Physicists are falling out over a high-profile claim that the speed at which gravity propagates closely matches the speed of light.

The row began after two researchers said that they had measured the 'speed of gravity' for the first time and found that, like the speed of light, it is finite.

Sergei Kopeikin, a theoretical physicist at the University of Missouri in Columbia, and Edward Fomalont, an astronomer at the National Radio Astronomy Observatory in Charlottesville, Virginia, revealed their results on 7 January, at the American Astronomical Society's annual meeting in Seattle.

The pair measured the refractive effects of Jupiter's gravitational field on the radio waves emitted by a bright, distant galaxy. From this they were able to determine that gravity propagates at close to the speed of light.

The value is similar to that predicted by Einstein, but several theorists were quick to assert that Fomalont and Kopeikin's interpretation of their results — which was widely reported around the world last week — is flawed.

"It's complete nonsense," says Peter van Nieuwenhuizen, a physicist at Stony Brook University in New York, who has devoted much of his career to studying gravity. Clifford Will, a physicist at Washington University in St Louis, Missouri, adds that he believes that the team's experimental set-up could theoretically yield gravity's speed of propagation. But Will's calculations suggest that the effect would be too small to detect with any existing telescope.

"The experiment is wonderful, but it has nothing to do with the speed of gravity," says Kenneth Nordtvedt, a retired physics professor at Montana State University in Bozeman. Nordtvedt says that the team is actually seeing a gravitational analogue of the force of magnetism, created by electrons moving at close to the speed of light.

"I think Dr Nordtvedt is confused," says Kopeikin. To get his speed-of-gravity interpretation, Kopeikin used his own simplified version of Einstein's gravity equations, which other theorists, such as Nordtvedt, have yet to fully accept. "The situation is a little upsetting for us," says Fomalont, who took the experimental measurements — adding that he hopes the quarrelling theorists will get together soon and settle their differences. ■

Safety doubts force rethink of embattled comet mission

Declan Butler, Paris

Rosetta, the US\$700-million European Space Agency (ESA) mission to land a probe on the comet Wirtanen in 2011, is set to miss its January take-off window and could be postponed for a year or more, *Nature* has learned.

The mission's postponement was set to be agreed on 14 January, after the failure of the Ariane 5 ECA 'heavy lifter' rocket on its maiden flight in December (see *Nature* **420**, 723; 2002) cast doubt on the safety of the standard Ariane 5 model, on which Rosetta was scheduled to fly.

Hopes that the launch would go ahead before the take-off window closes on 31 January were temporarily boosted last week, when an inquiry traced the cause of the accident to a cooling fault on the new engine — suggesting that the standard rocket is not at risk. And scientists at a pre-launch press briefing in London on 13 January said that they were still hoping for a January launch.

David Southwood, ESA's director of science, was scheduled to hold another briefing in Paris on 15 January "to take stock of the status of this mission". But *Nature* has learned that Arianespace, which operates the Ariane rockets, has advised against a launch — and ESA has agreed with its assessment. Southwood has previously said: "If there was the slightest risk in my mind, we would not



All dressed up, but nowhere to go: the Rosetta probe is set to miss its rendezvous with a comet.

launch; there is no way I'm not going to gamble a billion euros of other people's money."

The quickest alternative launch options would use gravitational assistance from Venus instead of Mars to swing the craft towards Wirtanen — but this would raise heat problems. Others would use Mars, as in the original plan, but this would mean finding a different comet to land on instead of Wirtanen. Sources close to ESA say that no plan will be finalized before a thorough analysis has been undertaken of the cost, scientific merit and technical risk of all options. ■

Report backs Smithsonian research

Erika Check, Washington

Researchers at the Smithsonian Institution (SI) in Washington DC are hoping that 2003 will mark a fresh beginning for science at the world's largest museum complex. That's because an independent panel has endorsed their work — and added that their main problems are a lack of leadership and money.

In a long-awaited report released on 7 January, the Smithsonian Science Commission says that research at the museums is "flourishing", but could "slip ... into a state of mediocrity from which it will be hard to recover" without more financial support. It says that the museums should try to secure more money from Congress, private sources and National Science Foundation grants.

The independent commission, chaired by Jeremy Sabloff, director of the University of Pennsylvania Museum of Archaeology and Anthropology in Philadelphia, says that aside from cash, lack of leadership and vacancies in key positions — such as the directorship of the National Museum of Nat-

ural History — have been "the most important factor in the weakening of SI science".

The Smithsonian's Board of Regents set up the commission in July 2001 after researchers publicly clashed with Smithsonian secretary Larry Small over several of his decisions (see *Nature* **410**, 727; 2001).

"We're glad that the report supports the value of science in the institution," says Bernard Finn, curator of electrical collections at the National Museum of American History and chair of the Smithsonian Congress of Scholars. "But the downside of stressing the need for more funding is that it allows the administration to say that we have to be content with losing staff and other resources, if they can't get more money."

Congressional staff members agree that it will be difficult to find more money for Smithsonian science amid a federal budget deficit and worries about an impending war in Iraq. But they say that the report's positive assessment of science at the Smithsonian should help its case. ■



Would-be cloners: Severino Antinori (left), Panayiotis Zavos and the Raelians' Brigitte Boisselier.

Prospect of human cloning poses dilemma for journals

Helen Pearson, New York

This time, it seems likely to have been a bizarre publicity stunt. But the media circus that has greeted the Raelian sect's claim to have produced two cloned babies raises an ethical dilemma. Should the scientific community help to verify any future similar claims? Or would this be condoning what most biologists agree is an unethical practice?

Leading scientists say that they would feel duty-bound to examine any serious claim that a human had been cloned. Perhaps more surprisingly, *Nature's* enquiries suggest that a success in human reproductive cloning would probably be published in a mainstream scientific journal — a move that many would interpret as an endorsement of the work.

Speculation about human clones won't go away. Italian fertility doctor Severino Antinori has claimed that one of his patients will give birth to a clone within weeks. And his one-time partner Panayiotis Zavos, based in Kentucky, asserts that he has women volunteers who are ready to carry embryos created by cloning.

Although many scientists view Antinori and Zavos as scarcely more credible than the Raelians, the birth of a human clone remains conceivable. "It's been possible for at least five years," says Tony Perry of the RIKEN Center for Developmental Biology in Kobe, Japan.

But would scientists who helped to corroborate a claim of human cloning be perceived as condoning the process? "By contributing to the debate, the inference could be that they're supporting it," says Chris Barratt, a specialist in reproductive medicine at the University of Birmingham, and a member of Britain's Human Fertilisation and Embryology Authority.

Most cloning researchers, however, argue that the public has a right to know the truth. "We have to become involved, even if we find the job repugnant," says Randall Prather of the University of Missouri-Columbia.

Prather is one of three cloning researchers who, in response to the Raelians' claims, released a statement urging them to submit to tests by a respected authority such as the US National Academy of Sciences. Britain's Royal Society, meanwhile, advocated that two independent, accredited labs should compare the DNA of the individual alleged to have been cloned with that of their 'offspring'.

For his part, Zavos says that he plans to follow established scientific protocol. "I intend to publish in scientific journals," he says. But would any journal agree to publish a paper describing work that animal data suggest would be associated with a high risk of miscarriage, birth defects and other health problems?

Nature requires work involving human subjects to have been approved by an ethical review panel and to meet the Nuremberg Code, the international guidelines for human experimentation. On these grounds, says Natalie DeWitt, the *Nature* editor who handles papers on cloning, a manuscript on human reproductive cloning would probably be rejected without peer review. Jeffrey Drazen, editor-in-chief of *The New England Journal of Medicine*, says that its editors make their own ethical evaluations, taking into account factors such as standards of informed consent.

But at least one journal would be prepared to publish such a paper if it passed scientific review. "If the reviewers thought it was OK, I wouldn't have any qualms," says Alan DeCherney, editor of *Fertility and Sterility*, published by the American Society for Reproductive Medicine.

Gerald Schatten, a developmental biologist at the University of Pittsburgh, Pennsylvania, says journals that publish such work may stand accused of condoning unethical experiments. "Even if one human clone was born in a normal fashion, it doesn't mean future experiments wouldn't have catastrophic consequences," he says. ■

German researchers set to receive Israeli stem-cell shipment

Haim Watzman, Jerusalem

For a team of scientists in Germany, the long wait is over. Collaborators in Israel this month began shipping them the human embryonic stem cells they need for their research.

Neuroscientist Oliver Brüstle and his group at the University of Bonn will receive the cells as part of a joint project with an Israeli team led by Joseph Itskovitz-Eldor of the Rambam Medical Center in Haifa.

Brüstle's team will investigate the way in which the stem cells develop into nerve cells. Although the work was approved by Germany's main research-funding agency, the DFG, last January (see *Nature* 415, 566; 2002), the grant only came through last month.

German law forbids the production of human embryonic stem-cell lines, and research on such cells can be performed only on imported cell lines that were derived before January 2002.

By contrast, Israel is one of the few countries — including South Korea, Sweden, Singapore and the United Kingdom — to permit the production and propagation of human embryonic stem cells, as well as research on them.

Brüstle says he feels that Germany's rules are too strict, as they mean that the country's researchers will not have access to newly derived, and potentially more attractive, cell lines. "I very much hope that, as we gain more data, there will be increasing pressure to open up and even develop new cell lines," he says.

In Israel, an advisory committee on bioethics said in 2001 that Israeli scientists should be allowed to use cells from surplus embryos from *in vitro* fertilization treatment and from fetuses aborted up to nine weeks after fertilization, provided that the parents give informed consent.

But Amos Shapira, a law professor at



Oliver Brüstle is getting his stem cells from Israel.

Tel Aviv University and a member of the committee, says he worries that Israel may end up as a major exporter of embryonic cells to countries with stricter rules. If care is not taken, he warns, commercial and research motives could cause violations of the requirements laid down by the committee, such as informed consent. ■

Paris university blasted over Israel motion

Declan Butler, Paris

The governing board of France's largest research university is facing a hail of protest from researchers both at home and abroad over its demand that the European Union (EU) sever its most important formal scientific link with Israel.

The outcry is a response to a motion passed by the board of the University of Pierre and Marie Curie — also known as the University of Paris 6 — calling for Israel's participation in the EU Framework Programme for research to be suspended. Around 2,500 protesters demonstrated at the university's Jussieu campus on 6 January, and an Internet petition against the move has attracted some 24,000 supporters.

The university's 60-strong board passed a motion on 16 December saying that Israel should be excluded from the programme until Palestinian universities, which are in dire straits because of the Israeli occupation (see *Nature* 417, 209–210; 2002), are able to operate normally.

The motion, which was backed by trade unions and passed by 22 votes to 4, stopped short of its initial demand that the university itself also suspend research collaboration with Israel. It also called on the university and Israeli researchers to act to alleviate the plight of Palestinian researchers.

Among those backing the Internet petition are nine Nobel prizewinners — among them David Baltimore, Claude Cohen-Tannoudji,



Out in the cold: protesters condemn a proposal to exclude Israel from research collaborations.

Elie Weisel and François Jacob. Cohen-Tannoudji, an honorary professor at the university, wrote in the French newspaper *Le Monde* of his “shame” at the motion's contents.

Critics of the motion include the national council of university vice-chancellors and Luc Ferry, the education minister. They say that the idea of punishing Israel's universities runs counter to the values of scientific collaboration, and that research is one of the few areas in which Jews and Arabs can still interact.

The petition's instigator, Bernard Maro, director of the developmental biology laboratory at the Paris university, wants the board to revoke the motion at its next meeting on 27 January. Otherwise, he says that he will leave the university and take his lab and its 120 staff elsewhere.

Maro is not alone: last week he led a delega-

tion of scientists that presented Gilbert Bereziat, the vice-chancellor of the university, with a letter signed by 91 of its leading scientists, calling on the university to “denounce without ambiguity the motion, and apologize to our Israeli colleagues”.

Bereziat, a biologist, says that the policy of the university — which has strong links with Israeli scientists — has not changed as a result of the motion, and that he would never cut off contacts with Israeli scientists.

The board itself remains deeply divided on the issue. Michèle Glass-Maujean, a physicist and board member who was abroad during the December meeting, says that the decision is beyond the board's remit. “I was elected to help run a university, not to take political positions — particularly on a situation as delicate as this,” she says, adding: “The board would have done better to have done nothing.” Supporters of the motion complain of hate mail calling them “fascists”.

The vice-chancellor of the University of Paris 7, also on the Jussieu campus, intervened last week to block discussion of a similar motion by its board. Instead, the board passed one that lauds collaboration as fundamental to the international scientific community and a weapon against “extremism”.

Meanwhile, 17 vice-chancellors of universities in the Paris area took this idea a step further, by calling on the EU to extend its agreement with Israel to explicitly include Palestinian universities. ■

Pesticide firms ask to use human data to assess safety

Tony Reichhardt, Washington

A lawsuit due to reach court in March could settle a contentious stand-off between pesticide manufacturers and the US Environmental Protection Agency (EPA) over whether human test data can be used to help set legal limits for pesticides released into the environment.

A group of pesticide suppliers and CropLife America, which lobbies for the pesticide industry, are suing the EPA in the District of Columbia Circuit of the US Court of Appeals, after the agency issued a ban on accepting human test data.

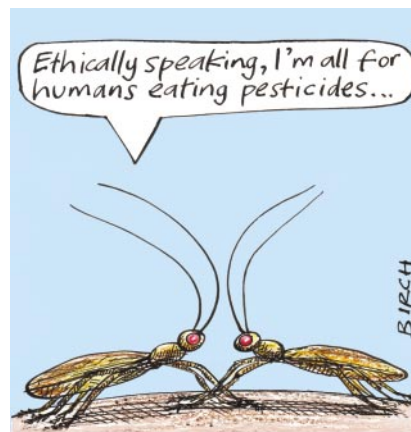
The outcome could affect sales of some of the mostly widely used pesticides on the market, including malathion, a crop insecticide, and dichlorvos, which is commonly found in domestic pest strips and flea collars.

Although such chemicals have been used for decades, a 1996 law tightened regulations governing their presence as

residue in food (see *Nature* 396, 207; 1998). Because their toxic effect on humans is often unknown, the EPA takes the lowest toxic dose based on animal studies and divides by 100 to set limits for humans.

Pesticide manufacturers say that the uncertainty factor is too conservative in many cases, and that using data from tests on humans would allow the limits to be raised for some pesticides. With this in mind, they have submitted human toxicity data for at least 15 chemicals since the 1996 law came into effect.

But a 1998 report by the Environmental Working Group, a Washington-based lobby group that wants tighter controls on pesticide use, exposed details of human tests in Scotland (see *Nature* 394, 515; 1998), and raised ethical concerns about the practice. The EPA subsequently banned the consideration of human data in setting pesticide regulations, a decision it reaffirmed in December 2001, when it asked



the National Academy of Sciences to look into the matter.

The academy panel charged with doing this, co-chaired by James Childress, a University of Virginia bioethicist, and Michael Taylor, a food-safety expert with

Ethics panel attacks environment book

Alison Abbott, Munich

The author of the *The Skeptical Environmentalist*, a widely read book arguing that the global environment is basically in good shape, misused scientific data to support his arguments, the Danish Committees on Scientific Dishonesty has ruled.

The book was written by political scientist Bjørn Lomborg, who now heads Denmark's Institute for Environmental Assessment (IEA). In an unusual judgement issued on 6 January, the committees said that the book was, "objectively speaking, deemed to fall within the concept of scientific dishonesty". The report adds, however, that the book cannot be regarded as scientifically dishonest, because there is no evidence that Lomborg was grossly negligent or intended to deceive. It concludes instead that the book's publication "is deemed clearly contrary to the standards of good scientific practice".

The verdict was welcomed by researchers who have accused Lomborg of selective use of data and of falsely presenting his work as a formal scientific analysis. But others said they were astonished that a scientific ethics panel would pass judgement on a polemic — even one published by Cambridge University Press with thousands of references and footnotes.

Anthony Trewavas, a plant biologist at the University of Edinburgh, says that, although he doesn't always agree with Lomborg, the move "comes down almost to censorship", and that the committees should



Bjørn Lomborg contends that an ethics committee was the wrong forum to judge his book.

never have agreed to consider the case in terms of scientific dishonesty: "Everyone has a right to express their opinion," he adds.

"There was not even unanimity within the committees or the working group set up to study the case as to whether it was in our sphere," Hans Henrik Brydensholt, the committees' chairman, told *Nature*. "But several issues convinced us that we could not choose simply not to comment."

For example, Lomborg identifies himself in the book as a scientist, and the University of Aarhus, where he is an assistant professor of statistics in the political-science depart-

ment, listed it as a research monograph in its 2001 yearbook. Moreover, the 500-page book is set out like a scientific treatise; it includes almost 3,000 notes, some 2,000 references to scientific papers, and was published by an academic publisher. "Most importantly," explains Brydensholt, a retired high-court judge, "it has been discussed widely in the scientific community."

Support for the Danish committees came from Hans-Heinrich Trute, the ombudsman for Germany's main research funding agency, the DFG. He says that his office, which considers misconduct cases in Germany, would have authorized an investigation in such circumstances.

The Committees on Scientific Dishonesty was also criticized by Lomborg and his supporters for relying on already available critiques of the book, such as those published in *Scientific American* in January 2002, rather than conducting its own analysis. "We thought long and hard about whether to institute an international panel of experts to reanalyse the issues," says Brydensholt, "but decided that it was not necessary."

Lomborg strongly contests the verdict. "The committees made no judgement on the substance of the book, only the methodology," he complains. He argues that as a policy book it should not in any case have been judged by a scientific committee. "It is a popular book aimed at a broad audience," he says.

The committees' findings seem unlikely to threaten his position in Denmark. Ole Christiansen, an economist and chairman of the IEA's board of directors, says that, as the committees did not raise any new accusations against Lomborg, he would recommend that the board take no action against him. "The real debate took place a year ago," Christiansen says.

But the IEA may yet face pressure over Lomborg's position. The institute was created by a right-leaning coalition government that ousted the long-ruling Social Democrats in the November 2001 elections. Now the opposition parties say the committees' verdict confirms their criticisms of Lomborg's appointment, which they say was politically motivated. They have already called for his removal.

The government responded by saying it would consider commissioning an international evaluation of the eight or so reports that the IEA has so far produced.

On the international front, several scientists praised the committees' verdict. "I am relieved that an independent body has clarified to non-scientists that the book is an opinion piece, not a rigorous scientific work," says David Tilman, a biodiversity expert at the University of Minnesota.

Resources for the Future, held its first public meeting in Washington on 8 January, and plans to report its findings in December.

Industry representatives claimed at the meeting that the EPA's ban on human data violates its legal obligation to consider all relevant data in making pesticide decisions. California and the European Union each consider human test data in setting pesticide standards, they said.

A study funded by CropLife America that was presented at the meeting found that all 15 of the oral human tests submitted as evidence to the EPA since 1996 complied with accepted standards regarding informed consent, institutional review and other ethical practices.

Steven Lamm, an occupational-health consultant in Washington, said that "excluding ethically performed human studies is unethical".

But Jennifer Sass, a toxicologist at the Natural Resources Defense Council, an

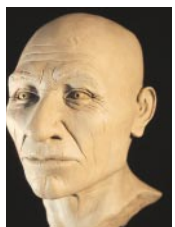
environmental lobby group, said at the meeting that her review of several human pesticide studies revealed flaws such as inadequate sample size or focusing on a single end-point in determining toxicity while ignoring other adverse health effects.

In theory, the EPA should be able to reject flawed studies, but it lacks the expertise to evaluate large numbers of human clinical studies, says Lynn Goldman, an environmental-health scientist at the Johns Hopkins Bloomberg School of Public Health and former head of the EPA's pesticides office.

Richard Wiles, vice-president for research at the Environmental Working Group, says that the EPA is under "enormous pressure from industry" to lift the ban on human tests. He adds that even if such studies are flawed, pesticide manufacturers will use them to try to influence debate about what constitutes a "safe" level of exposure to common pesticides.

Court rules in favour of research on Kennewick Man

San Diego After some 9,300 years in the ground, and a further six in storage at a museum, the skeleton of Kennewick Man could finally be made available to researchers in a few weeks' time. A federal court ruling on 8 January seems to have cleared the way for the remains to be studied, leaving anthropologists and archaeologists hopeful of beginning work as soon as April.



Native Americans want to bury Kennewick Man.

Kennewick Man was found in a riverbed in Washington state in 1996. It is one of the earliest hominid specimens found in North America, but has only briefly been examined. Further research has been delayed by a long-running legal dispute between researchers and

the US Department of the Interior, which argued on behalf of native American tribes that they should be allowed simply to bury the skeleton.

Judge John Jelderks said that the scientists should be allowed to examine the bones. This ruling follows an appeal by the tribes against a separate decision last year to allow the skeleton to be studied (see *Nature* **419**, 870; 2002).

The scientists are now in discussion with government officials on precisely when the research can begin. Tribal officials still hope to block the research in the courts, but observers of the case say that further delays are unlikely.

UK wants to swap facts for issues to teach bored teens

London The British government is considering reforms that would see many teenagers abandon traditional school lessons in physics, chemistry and biology in favour of studies of contemporary scientific issues.

The proposed changes in the national curriculum for children aged 14–16 include a greater emphasis on issues such as genetic modification, the limitations of scientific data, and discussions on media coverage of science. Lessons in conventional core science subjects would be scaled down to accommodate the changes, which are largely aimed at students who will not study science or engineering at university; supplementary courses in traditional subjects will be retained for those who intend to do so.

Some education groups have criticized the plans, which are likely to be included in a government discussion paper to be released

later this month. But others argue that they are needed to address waning interest in these subjects. A parliamentary committee last year criticized science lessons as so “boring” and “overloaded with facts” that they put people off science for life (see *Nature* **418**, 266; 2002).

Senators turn up heat on greenhouse gases

Washington Two prominent US senators have introduced legislation that would set up a trading regime for major producers of greenhouse gases, as well as setting overall targets for US greenhouse-gas emissions.

The bill, jointly sponsored by senators Joe Lieberman (Democrat, Connecticut) and John McCain (Republican, Arizona), was introduced at a hearing of the Senate commerce committee on 7 January. Its introduction by two of President George W. Bush's main political rivals as first order of business in the new Congress is viewed as a direct challenge to what critics see as the Bush administration's inertia on the climate-change issue (see page 195).

The Lieberman–McCain bill would seek to stabilize US carbon emissions at 2000 levels by 2010, and cut them to 1990 levels by 2016. It would also accelerate climate-change research and establish a scholarship programme at the National Science Foundation for young scientists in the field. But the measure faces opposition in the Senate and probably has even less support in the House of Representatives and the Bush administration.

University of California faces wave of budget cuts

San Diego Students heading for the University of California this autumn look set to face hefty fee increases as the university responds to a fresh cut of \$300 million in its annual budget.

The university, which runs nine campuses around the state, learned the bad news when details of the proposed state budget were released by governor Gray Davis on 10 January. The proposal contained some relief — no new cuts were suggested in the university's state-funded research programme. But sharp cuts of some \$80 million in the \$300 million that the state spends each year on research in the university had already been announced (see *Nature* **421**, 5; 2003).

The proposed \$96.4-billion state budget will now be considered by state legislators and the details finalized in the summer. In response, observers say, the university will probably introduce a range of other austerity measures, perhaps including a freeze on new appointments and restrictions on travel.

Noise protests drown out whale sonar tests

San Francisco Trials of a new sonar intended to protect whales have been suspended, because of fears that the device itself could damage the animals. Biologists at the Woods Hole Oceanographic Institution in Massachusetts say their sonar could alert sea-users, such as large ships and oil-exploration companies, that whales are in the area.

The researchers wanted to test the sonar off the coast of California earlier this month. But environmentalists said it should not be used because the noise it generates, which is louder than a jet engine, could injure migrating grey whales (*Eschrichtius robustus*). They say the situation is not serious enough to justify the risk. On 8 January a federal court agreed, and suspended tests pending further enquiries.

Underwater sonar has been linked to whale deaths before (see *Nature* **415**, 106; 2002), but the biologists insist that their equipment is safe as it operates outside the hearing range of the whales. A final decision is expected on 17 January.

Avian botulism clips wings of Taiwanese spoonbills

Taipei Conservationists and researchers in Taiwan are investigating an outbreak of avian botulism that has wiped out up to 10% of the global population of an endangered bird in the past two months.

Over 70 black-faced spoonbills, *Platalea minor*, at the Chiku Lagoon have succumbed to the disease. Current estimates put the worldwide population of the birds as low as 700. The birds breed on the eastern coast of the Korean peninsula and spend winter in Taiwan, China and Vietnam.

Warm winter weather, which encourages the growth of the bacteria that produce the botulism toxin, is being blamed for the outbreak. Researchers from Japan and South Korea met their Taiwanese counterparts this month to discuss the problem. Some suggested attaching transmitters to 12 spoonbills at disease-control centres, to track their feeding habits, but critics say this could stop the birds flying normally. Others say the answer lies in better wetland preservation.



Endangered: 10% of the world's black-faced spoonbills may have died of avian botulism.

Burning issues

Australia's cities impinge upon an ancient landscape shaped by fire. Carina Dennis talks to the researchers who are striving to protect lives and property, while retaining natural fire regimes that nurture the country's biodiversity.

Searing winds, soaring temperatures and smouldering homes — it's summer in Sydney and the suburbs are besieged. In recent months, Australia's largest city has been hemmed by wildfires. Highways have been closed, power lines severed, and scores of houses destroyed.

The need to protect urban areas from incineration is just one reason why fire science is a hot issue in Australia. Much of the country's unique flora and fauna has evolved with fire, coming to depend on the bush periodically going up in flames. But across large tracts of the country, established rhythms of fire and regrowth have been altered by human activities, in some cases stoking uncontrollable conflagrations that threaten lives, property and biodiversity.

Fighting fire with fire may offer the best solution. Controlled fires can be used to reduce the build-up of flammable vegetation, and might also mimic the natural cycles of scorching and regrowth needed to maintain biodiversity. But if fires are to become an effective land-management and conservation tool, scientists must improve their understanding of how complex fire regimes have shaped Australia's ecosystems. "We need a good understanding of how fire behaves in the landscape," says Jim Gould, a bushfire expert with the Forestry and Forest Products Division of the CSIRO, Australia's main national research agency, in Canberra.

But the residents of Sydney and other Australian cities have more pressing concerns. As cities have expanded into bushland, more and more urban dwellers have found themselves living in an extremely vulnerable position. "In Sydney, about 300,000 hectares lie at

this urban–bush boundary," says Ross Bradstock, a fire ecologist with the New South Wales National Parks and Wildlife Service.

Recent decades have unleashed some ferocious blazes — 16 February 1983 is remembered as 'Ash Wednesday', after fires swept through the states of Victoria and South Australia, killing 76 people and destroying more than 2,500 homes. During the fires of 1994's dry season in Sydney, four died and about 180 houses were lost. And the fires that threatened Sydney in December 2001 destroyed more than 100 homes.

Time bomb

The Australian bush — dominated by eucalyptus trees — is highly flammable. As the summer wears on, the build-up of undergrowth creates a natural tinderbox. The current drought, which dates from March 2002 and is the worst on record, has exacerbated this volatile situation. Add in the unpredictable variable of twisted individuals who get their kicks from deliberately starting fires, and the suburbs of Sydney and other Australian cities are like incendiary bombs with very short fuses.

With the public demanding action, federal and state governments have increased investment in firefighting equipment and stepped up surveillance. Science has also been called to the frontline, with the announcement in December of a new national Bushfire Cooperative Research Centre, based in Melbourne, which will bring together fire researchers from across the country in a seven-year, A\$112-million (US\$64-million) programme to investigate better ways of preventing and controlling bushfires,

particularly at the urban–rural interface¹.

The centre's main focus will be to support projects directly relevant to those responsible for preventing and fighting fires. For instance, some of the participating scientists have previously revealed how fires can suddenly threaten firefighters' lives following a change of wind direction. A fire that has been advancing for some time in the same direction tends to move forwards in a 'V' shape, with an intense, narrow 'head fire' at the apex. But even a slight shift in wind direction creates a wider, linear head fire. And by analysing fire behaviour under various conditions, CSIRO researchers have found that these linear head fires travel up to three times faster than was previously thought, reaching speeds of up to five kilometres per hour². "This is critical information for good safety training of firefighters," says Gould.

If towering flames that can change direction at a moment's notice weren't enough to contend with, combating bushfires in Australia is made even more difficult by the phenomenon of 'spotting'. Spotfires occur when firebrands of bark are tossed into the updraft of hot air from a fire and are carried ahead of the fire front by the wind. Although other countries have problems with spotfires, the highly flammable bark of many Australian trees makes spotting a particular hazard.

The CSIRO's Forestry and Forest Products Division has developed models of firebrand behaviour to predict spotting distances from different sources of burning bark, under various weather conditions. Using a specially designed wind tunnel³, the researchers record the 'terminal velocity' — the speed at which the firebrand begins to fall to the ground once



The threat to lives and wide-ranging destruction caused by Sydney's bushfires have led researchers to seek fresh ways to beat the blazes.



away from an updraft — monitoring the effects of firebrand size and length of time it was exposed to flame. In an initial unpublished study, Peter Ellis has concentrated on the messmate stringybark eucalyptus (*Eucalyptus obliqua*), which is common in southeast Australia and can shoot out huge numbers of firebrands that can travel up to five kilometres. “We plan to develop a database on the behaviour of a number of different bark types,” says Ellis.

CSIRO researchers are also recording the distribution of spotfires in high-intensity experimental fires during summer conditions. The eventual goal is to develop a comprehensive spotting model to help firefighters anticipate bushfire behaviour and assist communities in managing vegetation around residential boundaries.

Cutting through the smoke

Modelling fire behaviour is crucial, experts agree. “Fire research needs to go from being a descriptive to a quantitative science,” says Michael Reeder, a mathematician at Monash University in Clayton, near Melbourne, who is not part of the new cooperative research centre. Working with a group at the US National Center for Atmospheric Research in Boulder, Colorado, Reeder’s team has built a computer model of how a fire behaves under different conditions, including variable topographical and meteorological factors, such as strong winds or high temperatures.

In unpublished work, the researchers have tested their model’s predictions against experimental fires filmed with an infrared camera. “We have found that our model agrees with how the fire behaved,” says

Reeder. But he cautions that the model has so far only been tested for a controlled fire in a lightly grassed, disused airfield. Reeder eventually plans to simulate fire behaviour under more extreme conditions, which would be of most use to firefighters.

Research on extreme fires has determined the maximum blaze intensity that can be safely controlled by ground-based firefighters⁴. Bushfire intensity is characterized by the amount of energy released at the front of the main fire. “The maximum intensity of a directly controllable bushfire is about 3,500 kilowatts per metre,” says Gould. Some of the blazes that threaten Australia’s cities have intensities of tens of thousands of kilowatts per metre, and can only be held in check by bulldozing firebreaks into the landscape, or by deploying aircraft that can release large volumes of water.

To prevent fires from reaching such intensities, land managers can use prescribed burns early in the season to stop the build-up of dried undergrowth that can fuel an uncontrollable blaze. But the question of how much to burn is a huge bone of contention. Many city dwellers want to increase the amount of prescribed burning to protect their homes. But conservationists argue that burning too frequently can compromise biodiversity, killing the seeds of plants before they have a chance to germinate, and destroying wildlife habitats. A recent study across the state of Victoria suggests that biodiversity is currently suffering from too much fire around settlements and too little in many forests and national parks⁵. “We need to find a balance between maximizing biodiversity and minimizing risk to people and property,” says Kevin Tolhurst, a fire ecologist based at the

University of Melbourne’s Creswick campus, northwest of the city.

Finding the right pattern of burning to support Australian ecosystems is far from simple. “Some species like a lot of fire and some species don’t like much at all,” says Bradstock, who is fascinated by how distinct ecosystems with very different fire requirements can live side by side. For instance, in the Blue Mountains National Park, west of Sydney, the dry eucalyptus forests that crown ridges and blanket escarpments are regularly singed by bushfires. Yet deep in the gullies, rainforests — including the last refuge of the prehistoric Wollemi pine (*Wollemi nobilis*) — survive by only rarely encountering fire. Modelling the key components of fire and ecology, Bradstock has developed computer simulations to devise appropriate fire regimes for different ecosystems⁶. Under the cooperative research centre, he plans to refine the model by studying the influence of climate, weather and fuel load on fire ecology.

Patched up

Optimal fire regimes won’t simply involve prescriptions of fire frequency and intensity — burning needs to be patchy. “You need a balance of young and old vegetation,” says Tolhurst. Some Australian animals, such as the pebble-mound mouse (*Pseudomys chapmani*) and the ground parrot (*Pezoporus wallicus*), live on the edges between burnt and unburnt vegetation, seeking food from areas recovering from fire but raising their young in the shelter of unburnt older vegetation.

Tolhurst has just completed an unpublished study of the effects of five different fire treatments across about 50,000 hectares in



Heated debate: researchers such as Jim Gould (above) are desperately trying to find out more about fire's behaviour. Among other questions, they want to investigate the value of controlled burning (inset).

the Wombat State Forest, some 80 kilometres northwest of Melbourne. The work revealed a complex network of relationships between tree growth, soil nutrients and the abundance of species of insects, birds and mammals immediately after a fire. Ecosystems generally recovered within two years, as long as they weren't entirely razed. "A key conclusion from our study was that recovery from fire is rapid provided there are small unburnt patches left behind," says Tolhurst.

The biggest challenge is quantifying the extent of patchiness needed, which Tolhurst hopes to overcome with computer modelling. But progress may be limited by what is known about the historical pattern of fire in the landscape. "Data on fire histories are sketchy at best," says Malcolm Gill, one of Australia's foremost fire ecologists, who recently retired from the CSIRO's Plant Industry Division in Canberra. He believes that the answer is a national fire-mapping project to record the frequency, intensity and nature of bushfires across the country. This would include extensive satellite imaging and analysis of biological indicators, such as fire scars in trees. But Gill remains frustrated by "the lack of action so far" — and hopes that such an endeavour will eventually be supported by the new cooperative research centre.

Some fire ecologists, meanwhile, feel that the new centre should do more to address the issue of bushfires in northern Australia, where more than 30 million hectares of bush go up in flames every year⁷. "The fires in Sydney are a drop in the bucket compared with the area burnt every year in northern Australia," says Dick Williams, an ecologist with the CSIRO Sustainable Ecosystems Division in Darwin.

These conflagrations are thought to be one of the main culprits behind the declining biodiversity seen in the region, but have not attracted the same interest from politicians as the fires threatening southern cities.

Trailblazers

Williams is part of a team headed by Alan Andersen, also at the Sustainable Ecosystems Division, which has conducted one of the world's largest fire experiments to measure the impact of different fire regimes on entire ecosystems over some 250 square kilometres of northern Australia's tropical savanna⁸. Set up in 1988 at the Kapalga Research Station in Kakadu National Park, Northern Territory, the project has monitored the effects of four different fire regimes: absence of fire, annual fires early in the dry season, annual fires late in the dry season, and fires lit progressively through the season. The researchers have been crunching their data since the completion of the experimental phase in 1996. One of the most important findings is that fire frequency — rather than the intensity of individual fires — is the more important determinant of biodiversity. Put simply, too much of the bush is burning too often, destroying vital havens for small mammals and other species.

Worryingly for those trying to protect the region's unique flora and fauna, it seems that the frequency and intensity of bushfires in northern Australia has increased since European settlement. To many experts, an answer to the problem may lie with the Aboriginal people whose small-scale, deliberate burning shaped the landscape after their arrival in north Australia from Southeast Asia more

than 40,000 years ago. This 'fire-stick farming' was motivated by the Aboriginal people's own needs. But, in combination with climate change, it may have created a new equilibrium under which biodiversity flourished. Certainly, there is some evidence to support this view: one long-term study of an area with unbroken traditional fire management in Arnhem Land, some 250 kilometres east of Darwin, found that it had a good bill of health in terms of biodiversity⁹.

"Aborigines used fire to stabilize an incredibly flammable environment and very unstable situation," argues David Bowman, a landscape ecologist at the Northern Territory University in Darwin. The problem now, he suggests, is that European settlement and agricultural practices have disturbed that equilibrium — returning Australian ecosystems to a volatile 'pre-human' state, in which destructive late-season fires are more common.

Some researchers are now collaborating with Aboriginal groups — which, following successful land-rights claims, own a large proportion of northern Australia's savanna — to reintroduce mosaics of deliberate burning in some areas. But they are working in a political minefield. Australia's racial politics remain fraught, and some cattle ranchers argue that the burning threatens their livelihoods.

Other experts, in any case, are less convinced that the answer to Australia's current bushfire problems necessarily lie with a return to the practices of the people who first managed the bush by fighting fire with fire. "Any blanket statement that Aboriginal burning is best for biodiversity is more ideological than scientific," says Andersen.

But the country's fire scientists agree that the interests of city dwellers and biodiversity alike require a much more comprehensive and quantitative approach to studying the blazes that have shaped the Australian landscape — with the new cooperative research centre representing just the start. "We have to learn to live with fire, whether it be from an economical and social point of view or from an environmental perspective," says Tolhurst. ■

Carina Dennis is Nature's Australasian correspondent.

1. Dennis, C. *Nature* **420**, 723 (2002).
2. Cheney, N. P., Gould, J. S. & McCaw, L. *Aust. For.* **64**, 45–50 (2001).
3. Knight, I. K. *Fire Technol.* **37**, 87–100 (2001).
4. Loane, I. T. & Gould, J. S. *Aerial Suppression of Bushfires: Cost/Benefit Study for Victoria* (National Bushfire Research Unit, CSIRO Division of Forest Research, Canberra, 1986).
5. Fire Ecology Working Group *Analysis of Disturbance by Fire on Public Land in Victoria* (State of Victoria Dept of Natural Resources and Environment & Parks, Victoria, Melbourne, 2002).
6. Bradstock, R. A. & Gill, A. M. *J. Mediterr. Ecol.* **2**, 179–195 (2001).
7. Williams, R. J., Griffiths, A. D. & Allan, G. in *Flammable Australia: The Fire Regimes and Biodiversity of a Continent* (eds Bradstock, R. A., Williams, J. E. & Gill, A. M.) 281–304 (Cambridge Univ. Press, Cambridge, 2001).
8. Andersen, A. N., Cook, G. D. & Williams, R. J. (eds) *Fire in Tropical Savannas: The Kapalga Experiment* (Springer, New York, in the press).
9. Yibarbuk, D. et al. *J. Biogeogr.* **28**, 325–344 (2001).

The times, they are a-changin'

More accurate timepieces could lead to better global positioning systems, insights into fundamental physics and a redefinition of the second. David Adam rates the runners in the race to build tomorrow's atomic clocks.

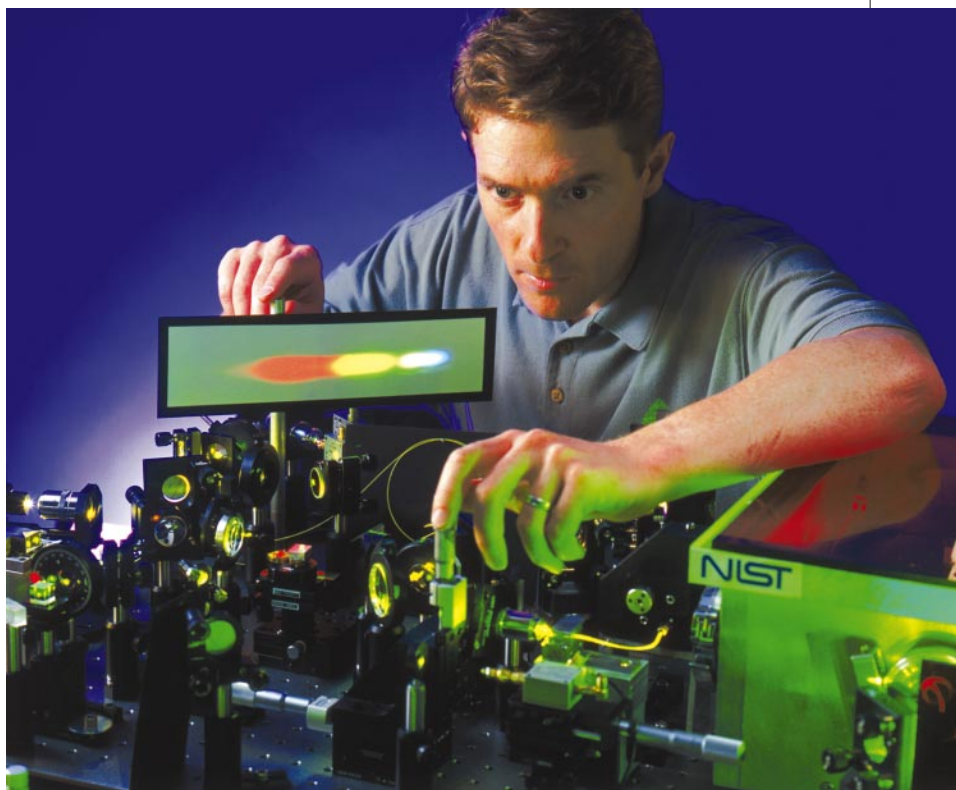
It's perhaps not the ultimate race against time, but rather the race to keep ultimate time. In a handful of labs around the world, physicists are developing a new generation of clocks so accurate that they should lose just one second in 100 billion years. By taking the pulse of laser light synchronized to the beating of atoms or ions, the researchers are confident that they can create clocks up to 1,000 times more accurate than today's best timepieces.

Such precision won't make much difference in the everyday world — trains will still run late and eggs will still take around four minutes to boil. But with the proposed optical clocks, studies of fundamental physical constants could become much more powerful. Global positioning systems (GPS) should also see a boost in precision, allowing them to pin down locations to within a few centimetres. What's more, if the optical clocks become established — and a prototype has already been built — they could force scientists to redefine the way we measure the basic unit of time itself: the second.

The push towards the new clocks is being led by the national standards laboratories in Britain, Canada, France, Germany and the United States. Each centre is developing its own particular design, but they all have one thing in common. At their hearts are the two basic components of every timepiece invented since the sundial: an oscillator, to provide events that repeat with a regular period, and a counter.

The most accurate existing timepieces, known as atomic clocks, measure a second using caesium atoms. By absorbing microwave photons of a particular frequency, these atoms move between two closely spaced energy states, emitting a photon as they return to the original state. Clock operators tune the frequency of the microwave radiation so that it resonates with the movement of the atoms between energy states, searching for resonance by monitoring the number of photons absorbed. The second is then defined as 9,192,631,770 cycles of the microwave radiation at this resonant frequency. The best caesium clocks can measure a second accurately to 15 decimal places.

Atoms and ions can also be excited to higher energy states by absorbing optical photons. And in the same way that a grand-



First timer: Scott Diddams' prototype mercury clock was the first timepiece to use optical wavelengths.

father clock with a pendulum that swings once a second is more accurate than a sundial, which repeats its cycle over 24 hours, so clocks that can count transitions driven by the higher frequencies of optical radiation are potentially much more accurate than microwave systems. In theory, an optical clock could measure a second down to a staggering 18 decimal places.

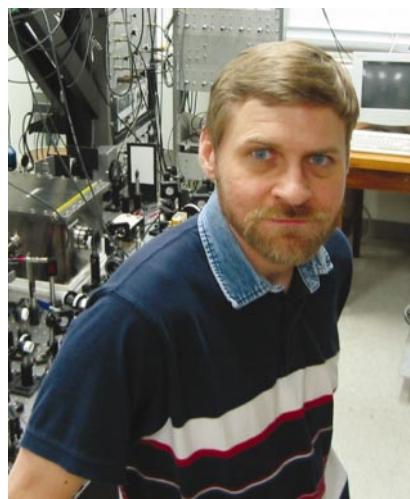
The current undisputed leaders in this effort are Scott Diddams and his colleagues at the National Institute of Standards and Technology (NIST) in Boulder, Colorado. In

In theory, an optical clock could measure a second down to 18 decimal places.

August 2001, they demonstrated a prototype optical clock for the first time (S. A. Diddams *et al. Science* **293**, 825–828; 2001). Diddams' clock harnesses the 10^{15} oscillations per second made by a laser that is exciting a single mercury ion. The way in which the ion absorbs the laser light is monitored, and the laser is adjusted so that it resonates with the ion's movements between energy levels. The researchers are currently trying to establish how accurate the clock is. "There's still a long road ahead to make real devices that can function day-in and day-out," says Diddams.

To improve matters, the NIST team is studying the factors that control the accuracy and stability of their clock. These are two separate issues — accuracy defines how closely a clock's output matches the desired time interval, whereas stability is a measure of how steady that output is. A clock that loses precisely a second each day is inaccurate but stable, for example.

G. WHEELER/NIST



Patrick Gill (left) and Alan Madej predict that optical clocks will rise to prominence within a decade.

The clock's height above sea level, for instance, influences its accuracy, as the rules of general relativity dictate that gravity affects observations of all timepieces — in this case the resonant frequency of the mercury ion. To achieve maximum accuracy, differences in height between the laser and the apparatus that counts the laser's oscillations need to be corrected for. Minute motions in the laser, on the other hand, can shift the frequency of the laser light and cause stability problems. Alan Madej of Canada's Institute for National Measurement Standards in Ottawa is working on an optical clock that uses a strontium ion as its pace-maker. He points out that stability can be compromised if the laser moves by as little as the diameter of a neutron.

Timely developments

None of these problems should be insurmountable, however, and Madej says that optical clocks that outperform the best existing microwave clocks will probably be developed and tested within a decade or so. Movements due to vibration can, for example, be reduced by mounting equipment on tables that float on cushions of air. And several groups are working on clocks based on millions of atoms, which resonate together when excited by laser light and produce stronger and more stable signals. Collisions between the atoms do, however, cause shifts in the tell-tale resonant frequency, and different arrangements of atoms produce different shifts. So what this design gains in stability, it may lose in accuracy.

If optical clocks can be made to work, GPS could be one of the first areas of technology to benefit. Currently, receivers calculate positions using the time signal from in-range satellites run by the US military, each of which contains a basic microwave atomic clock. At present, the system can measure locations to within tens of metres, or better if the receiver is within range of one of the ground stations, which use more

accurate clocks. Improving the accuracy of the satellites' clocks could enable receivers to measure their positions to within a centimetre.

Some physicists who are interested in fundamental constants are also in need of new clocks. In 1999, a team of astronomers led by John Webb of the University of New South Wales in Sydney provocatively suggested that the fine-structure constant, which determines the strength of electromagnetic forces, may not actually be constant (J. K. Webb, V. V. Flambaum, C. W. Churchill, M. J. Drinkwater and J. D. Barrow *Phys. Rev. Lett.* **82**, 884–887; 1999). Webb's team looked at light emitted by very bright objects known as quasars, which are thought to be distant galaxies powered by a supermassive central black hole. During the billions of years it takes for their light to arrive at the Earth, certain frequencies of this light are absorbed by interstellar dust.

But analysis of the quasars' light showed that the absorption, which depends on the value of the fine-structure constant, did not match that predicted from laboratory experiments. This, Webb suggests, may be due to the fact that the value of the constant has changed during the light's long journey — although many astronomers argue that his data are not sufficiently robust to make such a bold claim.

New optical clocks could help to test Webb's controversial idea. The value of the



Spot-on: could optical clocks improve GPS?

fine-structure constant influences the resonant frequency of ions, so researchers want to compare the beats of two clocks made from different ions. If the ratio changes over time, variations in the fine-structure constant could be the cause. Attempts to investigate this by analysing the time signal from conventional atomic clocks have so far not produced any positive results, but this could be because they cannot measure time accurately enough.

Second thoughts

With caesium clocks currently being used to define the second, the development of a new optical clock could also put pressure on the International Committee for Weights and Measures, the guardian of global measurement standards, to redefine this unit. "It's likely that in maybe a decade we'll be thinking of a redefinition of the second using an optical clock," says Patrick Gill, who works on optical clocks that use ytterbium or strontium ions at Britain's National Physical Laboratory in Teddington, near London. But with several different countries all pouring money and effort into their own designs, it would be a brave decision to pick one as a model system. "I don't think there will ever be anything to say that one clock is really head and shoulders above the others," argues Madej. "I have a feeling that, politically, the second will not be revised."

The availability of different designs will, however, help physicists to test their devices. Researchers currently have to assess the performance of optical clocks using the lower frequencies produced by caesium clocks. This isn't as contradictory as it sounds. Ted Hänsch, a physicist at the Max Planck Institute for Quantum Optics in Garching, Germany, has developed a system that multiplies the frequency of the microwave signal, producing a new signal with a frequency that is in the optical range (T. Udem, J. Reichert, R. Holzwarth and T. W. Hänsch *Phys. Rev. Lett.* **82**, 3568–3571; 1999), and which can be compared to the output of the optical clocks. The system works for now, as the accuracies of optical and microwave clocks are similar.

But this technique will fail when more accurate optical clocks are developed. The first such clock will, by definition, have no other device with which to calibrate it. "When an optical clock finally surpasses the accuracy of microwave clocks, the most rigorous way to verify that is to build a second one," Diddams says. And it may not stop there. If the second clock disagrees with the first, another will be needed. "If you have only one clock then you don't know what time it is, but if you have two then nobody knows what time it is," points out Diddams. "You actually need three clocks." Even when the race to develop the first accurate optical clock has been won and lost, the physicists involved won't be left with time on their hands. ■

David Adam is a news and features writer for *Nature*.

Journals: redundant publications are bad news

Publishing the same work twice is unethical and casts doubt on the integrity of research.

Sir — We have developed an electronic systematic search tool to estimate the amount of duplicate publications in the 70 ophthalmological journals listed by Medline. Our results show that there is a considerable number of duplicate publications. If this holds true for other disciplines, it is bad news for research.

For our survey, we matched the title and author(s) of each of the 22,433 articles published in the 70 journals between 1997 and 2000 using a duplicate-detection algorithm¹, and found that 13,967 pairs of articles give a matching score of 0.6 or more. Of these, we manually reviewed a random sample of 2,210. We found 60 genuinely 'duplicate' publications and estimate that 1.39% of the analysed articles are redundant. Because of the very restrictive selection process and the impracticality of detecting all duplicate publications, and because the estimated amount of duplicates increases with lower matching scores (Fig. 1), we regard this estimate to be the tip of an iceberg.

Of the 70 journals, 32 were victim to duplicate publication — 27 journals published the first paper and 26 the duplicate, on average 6.4 months later (standard deviation 4.7, range 0–21.3 months). We found no statistically significant difference between the average journal impact factor of the first (1.13) and the second journal in which the duplicate article was published (1.42) (Wilcoxon-signed ranks test $P > 0.1$). The analysed publications were by 210 authors, suggesting by extrapolation that a total of 1,092 authors could have been involved in redundant publication during the time period that we analysed. The scientific conclusions of the original and of the duplicate(s) were identical in 88.3% of cases; we found slight changes in 6.7%; and major changes (different results despite identical samples, or omission of patients) in 5% of cases.

Duplicate publications are unethical. They waste the time of unpaid, busy peer reviewers and of editors; inflate further the already over-extensive scientific literature; waste valuable production resources and journal pages; lead to flawed meta-analysis; exaggerate the significance of a particular set of findings; distort the academic reward system and copyright laws; and bring into question the integrity of medical research. Reproduction of data yields no benefit other than to the authors.

It is important that journal editors can trust their authors. Although many duplicate publications are discovered by

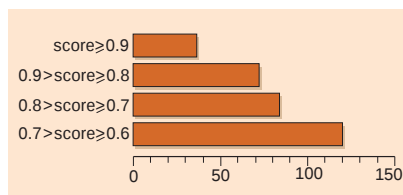


Figure 1 Estimated number of redundant publications for matching scores of 0.6 or more, where 1 = total overlap.

careful peer-reviewers or editors, they cannot provide complete protection. Scientific journals can combat redundant publication in various ways², but in practice the penalties for duplicate publication are minimal³.

Proper deterrents are needed: for example, better education on publication guidelines, the introduction of registers for planned and ongoing clinical trials, and a change in assessment criteria from quantity to quality when papers are submitted for posts or grants. As long as

Journals: how to decide what's worth publishing

Sir — Your News Feature (*Nature* **419**, 772–776; 2002) raises important questions about the reliability of peer review, but falls back on the justification often used by editors to shield themselves from widespread dissatisfaction with the system as currently practised: “If it ain't broke, don't try to fix it”.

We believe it may never have been working in the first place.

Perhaps peer review, in its current form, cannot be expected to detect fraud. But can we even rely on it to improve the chances that what is published is the best science, communicated as accurately as possible, and that what remains unpublished is dispensable?

Various studies, mostly in biomedical journals, have reported only modest author satisfaction (at best) with the review process, irrespective of the quality of the review. Papers that eventually became very highly cited were often rejected by the journal of first choice. Peer review is costly, biased, can be inefficient, does not always identify important work, and can allow publication of articles with serious deficiencies or omissions.

Rather than falling back on the churchillian cliché quoted in your feature that peer review is the worst system in the

publications remain the central requirement for academic advancement, a reasonable solution seems unlikely. Nevertheless, it is imperative that the problem of redundant publications be addressed, for it is the responsibility of all those who care about objective research and evidence-based medicine.

Stefania M. Mojon-Azzi*, **Xiaoyi Jiang†‡**, **Ulrich Wagner***, **Daniel S. Mojon‡§**

*Research Institute for Management in Health Services at the University of Applied Sciences, St Gallen, Switzerland

†Department of Electrical Engineering and Computer Science, Technical University of Berlin, Germany

‡Department of Ophthalmology, Kantonsspital, 9007 St Gallen, Switzerland

§Scientific Secretary, Swiss Society of Ophthalmology, Kantonsspital, St Gallen, Switzerland

1. Jiang, X. & Mojon, D. S. in *Proc. 1st Int. Workshop New Developments in Digital Libraries* 79–88 (ICEIS, Setúbal, Portugal, 2001).

2. Cho, B. K. et al. *Ann. Thorac. Surg.* **69**, 663 (2000).

3. Franken, E. A. *Acad. Radiol.* **5**, 407–408 (1998).

world except for all the others, members of the research community should cooperate to answer several questions.

We need to know whether peer review (in whatever form) is more effective than alternatives. Does it identify submissions of higher quality than do other selection methods, or chance, or no selection? Does peer review significantly improve the clarity, transparency, accuracy and usefulness of published papers compared with the submitted versions?

If peer review in its current, descriptive form is ineffective or less than effective, we should experiment with more analytical forms of assessment. For example, the quality of a new study could be assessed in the context of a pre-existing systematic review of studies on the topic. Such a population approach may make it easier to assess the contribution of an individual new study. At the same time, assessment should be standardized and specific for different experimental designs, and peer reviewers should be trained to use a single, structured-assessment instrument.

Ultimately, it is the larger population of readers (rather than a possibly biased sample of referees) who should decide whether the changes made during review substantially improve the document as a record of a peer's contribution to science. New systems should be tried that involve readers in the review process either after

'traditional' review (as in post-publication commentary, rapid replies and the like) or by developing a 'definitive' text by consensus before publication. Language experts have been investigating readers' reactions to texts for many years; it is time for editors and publishers in the 'harder' sciences to use their methods to extract useful experimental data from these reactions.

Tom Jefferson

Health Reviews Ltd, Via Adige 28a, Anguillara Sabazia, Rome, Italy

Karen Shashok

Comp. Ruiz Aznar 12, 2-A, 18008 Granada, Spain

Journals: impact factors are too highly valued

Sir — Linda Butler in Correspondence (*Nature* **419**, 877; 2002) shows that researchers in Australia are publishing more papers since the number of publications was introduced as a performance indicator for research. Butler points out there is now "little incentive to strive for placement in a prestigious journal. Whether a publication is a groundbreaking piece in *Nature* or a pedestrian piece in a low-impact journal, the rewards are identical".

The point is well-made, but her phrase highlights another growing problem in measuring performance which, if unchecked, threatens to have a major impact on science policy and progress. The problem is an over-reliance on journal impact factors to judge the worth of scientists.

It is increasingly common to hear scientists making snap judgements about the quality of others' work simply by perusing the names of the journals in which they publish, with no actual attempt to read their papers. This is a dangerous habit, for quite brilliant work can appear in a 'lesser' journal, either because its subject is not currently fashionable or because its author has special reasons for preferring a specialist forum. The habit is also dangerous because it erodes the capacity of the research community to determine its own direction.

An ex-colleague of mine, for example, liked to publish his excellent work on nerve regeneration, which could have been published anywhere, in a very specialist surgical journal because that is where he thought it would be most likely to inspire immediate clinical use.

The professional editorial staff of very high-impact journals such as *Nature* have a primary responsibility to the success of their journal: circulation, advertising, impact statistics and reputation. Deluged

with submissions from authors hopeful of publishing in a journal that will give them bench-credibility in a world of instant judgements, these editors must screen submitted papers to see if they meet the journal's needs before sending them out for peer review. Therefore, most submissions are rejected for reasons other than flawed scientific reasoning.

I have no criticism of this approach: it makes sense in the commercial world of journal production. The problem arises when scientists and administrators of science use the placement of papers to judge the worth of researchers, the worth of institutions, the best places to award grant money and the best places to fund fellowships. The more we couple allocation of resources to publication in 'top' journals, the more we are effectively handing over the direction of research to a small group of professional editors, who never sought this responsibility and who (excellent at their intended jobs though they may be) are unlikely to be the best people to bear it.

Most of us are, at least sometimes, the judges as well as the judged. If we do not consistently take the trouble to judge papers by their content rather than by their location, the direction of science will come to be determined, however unintentionally, by an editorial élite. We shall have only ourselves to blame.

Jamie Davies

Edinburgh University College of Medicine, Teviot Place, Edinburgh EH8 9AG, Scotland

Bright students enjoy correcting the textbooks

Sir — Students aged 16–17 have been doing chemistry research at Westminster School for the past five years (see the News feature "Put your lab in a different class", *Nature* **420**, 12–14; 2002). Our projects all have their origins in the normal curriculum, as many points of quite elementary chemistry have not been investigated for half a century or more. With modern techniques we can amplify (and often correct) what is written in the standard textbooks.

Our first paper, on the addition of hydrogen halides to alkenes, has now been published (*J. Chem. Soc., Perkin Trans. 2*, 810–813; 2002), and other work is nearing completion. Students gain by having to think about a problem for a year or more, and experiencing the disappointments as well as the satisfaction inherent in original work.

Too often, bright pupils are put off from studying science because they think they will be asked nothing more

demanding than to reproduce received wisdom. I hope that the presence of active research groups in high schools may help to correct this misconception.

Peter Hughes

Westminster School, 17 Dean's Yard, London SW1P 3PB, UK

Animal research needs organized defence

Sir — Your Opinion article "Promoting animal research" (*Nature* **420**, 447; 2002) delivers a much-needed message.

Ten years ago, I volunteered to join a National Institutes of Health programme to educate young people about the need for intact animals in biomedical research. Local high schools and colleges were sufficiently receptive to encourage me to continue.

In the past four or five years, however, my approach to the education authorities has fallen on deaf ears, although all of the teachers, and many of the students, voiced praise for the programme early on. One teacher told me that they did not want to run into problems with animal activists by allowing me to speak. Most of the population is disappointingly uneducated about science and a significant percentage is anti-science.

Next month, I will send approximately 30 letters to local colleges and high schools in an attempt to rekindle the interest. I believe that major biomedical and medical societies, and journals, should constantly urge an educational campaign to deliver our message to the public.

People like myself are very willing to volunteer to speak, design handouts, and so on, but a central focus group is needed.

Charles G. Smith

Address supplied

DNA discrepancy

Sir — We should be able to trust any author, whether or not a scientist, to deliver an accurate description of the past. Indeed, your final editorial of 2002 exhorts scientists to work to retain the public trust (*Nature* **420**, 719; 2002). Thus, it is even more unfortunate that *Naturejobs* states in the same issue (*Naturejobs* 3; 19/26 December 2001) that Rosalind Franklin and Maurice Wilkins worked on DNA structure at the University of Cambridge: they were famously, of course, at King's College London.

Alex May

Division of Mathematical Biology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

Emmet's inch and eagle's mile

How scientists sized up all they surveyed.

Measuring America: How an Untamed Wilderness Shaped the United States and Fulfilled the Promise of Democracy

by Andro Linklater

Walker/HarperCollins: 2002. 288 pp.

\$26/£17.99

The Measure of All Things: The Seven-Year Odyssey and Hidden Error that Transformed the World

by Ken Alder

Little, Brown: 2002. 422 pp. \$27, £15.99

Lewis Pyenson

Capturing the dimensions of the world has animated people since antiquity, but relentless precision is the hallmark of modern science, from Renaissance calendar reform to microelectronics and galactic astronomy. Commentators today associate faith in precision with a belief in progress, two prejudices increasingly absent in the post-modern world. *Measuring America* and *The Measure of All Things* remind us, however, that precision has always been a barely tolerated house guest of civilization, which prefers to tally its accomplishments and satisfactions as grandmother might around a holiday table, in subjective measure.

Measuring America provides an engaging perspective on how Americans legitimized their ownership of the land, divided up the terrain and sold it off. The process began in colonial times with English measures, notably the early-seventeenth-century Gunter's chain, the length and construction of which (100 links of 0.66 feet per link) allowed for an acre (10 square chains) to be laid out with ease. Andro Linklater shows that interest in land surveying stemmed from the notion of land as property, a significant development in late Tudor England. In the glow of early capitalism, buyers needed a precise accounting of their land, just as they reckoned a precise return on other commercial investments. In America, surveying remained the engine of capital accumulation until industrialization took over in the latter part of the nineteenth century.

The possession of land by ordinary people derives, as the nineteenth-century French anarchist Pierre-Joseph Proudhon might have affirmed, from theft. Henry VIII stole the lands of the Catholic Church; kings, governors, parliaments and congresses stole the lands of native Americans. Sale of crown or federal land became a major revenue stream. The property, often sold at a discount to insiders, brought fortune to speculators. Deeds drove before them those who had occupied the lands before the surveyors



Home on the range: when surveyors measured America it paved the way for private land ownership.

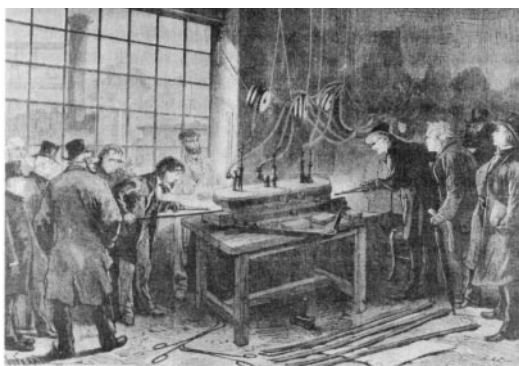
could do their work; these squatters raced west to claim a home in advance of the Guntermen, who legitimized the homestead. For a brief moment late in the eighteenth century, America had the chance to go metric (Thomas Jefferson was an advocate of it), but the urgency to survey and sell the lands of the Midwest seems to have doomed any reasonable alternative to emmet's inches and eagle's miles, as poet William Blake knew his measures in England. *Measuring America* provides a stunning picture of the pre-industrial origins of American wealth in circumstances that make current corporate scandals seem like children's games.

Increasingly precise measurements dominated the eighteenth century. In Europe, Antoine-Laurent Lavoisier's new chemistry emerged on the arms of unusually sensitive balances; the physics of Henry Cavendish, Charles-Augustin Coulomb and Jean-Charles de Borda were derived from careful instrumentation. The French Revolution was about nothing so much as grandness, and when the revolutionaries sought a standard for length based on something other than the king's foot, men of reason naturally attached their belief in precision to the world itself. They rejected a laboratory standard — the length of a pendulum oscillating in the time of one second — in favour of a project to survey France from top to bottom. The standard unit of length would become one

ten-millionth of a quadrant of a meridian measured from Dunkirk to Barcelona.

Along with major voyages of discovery in natural history, surveys were the largest scientific undertakings of the second part of the eighteenth century. From observing transits of the planet Venus at sites around the world to making a complete inventory of a river valley (Napoleon's of the Nile), the surveys had an appeal comparable to space exploration in the 1960s. French savants in the Academy of Sciences correctly sensed that a domestic survey would guarantee the survival of the exact sciences at a time of social change.

In *The Measure of All Things*, Ken Alder takes us through the meridian project in dramatic fashion, as the two principals, Jean-Baptiste-Joseph Delambre and Pierre-François-André Méchain, triangulated their way across revolution and war from the frontiers towards the centre of France. They hoped to obtain the meridian with unprecedented accuracy, using a new device for measuring angles, the Borda repeating circle, manufactured by the renovator of precision instruments in France, Etienne Lenoir. They slogged across the countryside with axemen to clear brush, carpenters to erect scaffolding, assistants to haul equipment, and servants for domestic chores, always looking for eyries and church steeples from which to sight distant markers. Much depended on the weather, on the season, and on the ability to



Setting the standard: scientists in Paris began constructing the metre in the 1870s.

commandeer resources or pay for them with Revolutionary notes of fluctuating value.

Rising from obscurity to become the protégé of Joseph-Jérôme Lalande, the first astronomer of the Enlightenment, Delambre took the northern two-thirds of the meridian; Méchain, also a social climber, took the southern part, including the difficult traverse of the Pyrenees. Early in the survey, Méchain took two incompatible sets of measurements for the position of Barcelona, an inconsistency he kept secret for his whole life, which ended when, after completing the seven-year project, he went back to Barcelona for another go at the end-point of the baseline.

Alder observes that the deception did not invalidate the standard metre (subsequent analysis of the data with Gauss-Legendre statistics, unavailable to the metric crusaders, seems to have lessened the gravity of Méchain's cover-up). One wonders if any of the regimes that eventually accepted the metric system in the nineteenth century cared about its naturalistic justification as much as its Revolutionary legitimization, and for that, any convention would have served as well.

If these fine books are an indication, story-telling has returned to the history of science. The narrative flows in lambent prose. Absent are theoretical caprice and postmodernist claptrap. We may be attending the birth of a scholarly enlightenment based on reason, clarity and elegance. ■

Lewis Pyenson is at the Center for Louisiana Studies, University of Louisiana at Lafayette, Lafayette, Louisiana 70504, USA.

Glue for the mental world

The Big Book of Concepts

by Gregory L. Murphy

MIT Press: 2002. 563 pp. \$45, £29.95

Paul Bloom

"Concepts are the glue that holds our mental world together." This is the opening statement of this superb book. Actually it is an understatement. Without concepts there would be no mental world in the first place. Concepts are mental representations that tie together specific instances, and are essential for relating present experiences to knowledge from the past. Concepts allow us to move from William James' "blooming buzzing confusion" to structured and adaptive thought. Without a concept of object, there would be no understanding of the physical world; without 'person' and 'friend', there would be no social cognition; without 'tomato' and 'apple', we would starve to death, staring dumbly at these things without any conception of what they can do for us. Without concepts there is no learning, and no innate knowledge.

The last time someone attempted an authoritative review of this field was in 1981, when Edward Smith and Douglas Medin published *Categories and Concepts* (Harvard University Press). They reviewed three general theories of what we have in our heads that makes categorization possible. The classical view proposed that concepts are necessary and sufficient conditions, or definitions. Hence membership is all-or-none: something either fits the definition or it doesn't; you cannot be a little bit pregnant.

The probabilistic view — also known as the prototype theory — comes in many flavours, but the most popular version is that concepts are lists of features, and that something belongs to the category to the extent that it possesses these features. This allows for category membership to be graded. Our concept of chair, for instance, allows for some things to be better (more prototypical) chairs than others, because some things match all of the features of the concept, whereas others, such as a beanbag chair, match just some of them.

Finally, the exemplar view posits that when we encounter objects in the world we record our individual experiences in memory, and that new objects are categorized to the extent that they are similar to other stored instances of that category. The radical idea here is that there is no summary representation. There is no single concept of chair, for instance; rather, something is a chair to the extent that it is similar to other objects that we have encoded as chairs.

After carefully reviewing the evidence at the time, Smith and Medin decided it was premature to choose between these theories. The evidence favours the probabilistic and exemplar views, but even so "there are aspects of the classical view that will doubtless find their way into any comprehensive theory of concepts".

The Big Book of Concepts shows us where we are some 20 years later. Murphy is one of the leading scholars in this area, and he reviews a complicated literature with honesty, clarity and wit. This is going to be the classic text in the field for a very long time. It is one of those rare cases in which the standard back-of-the-book blurb is actually true. Anyone seriously interested in concepts and categorization — seasoned researchers, graduates and advanced undergraduates, or scholars who simply want to get a sense of the field — must read this book.

What progress has been made since 1981? For one thing, a new theory of concepts has emerged: the knowledge approach, which posits that concepts are themselves embedded in rich conceptual structures, sometimes described as intuitive theories. This theory is supported by several empirical demonstrations — many by Murphy himself — that background knowledge can profoundly affect how concepts are learned, understood and used. There are also a lot more data. *The Big Book of Concepts* is so big because it reviews experimental research from several research areas that were barely discussed by Smith and Medin, such as the sorts of concepts formed by babies and young children, the comprehension of new conceptual combinations, and the effects of expertise on categorization.

If you skip to the end of the book looking for a clear resolution, you will be disappointed. Concepts, Murphy cheerfully concludes, are a mess. With the exception of the classical



Seat of learning? Chairs are recognizable as such because they are similar to other objects already encoded as chairs.

view ("a total flop"), each of the existing theories is the best explanation of a particular set of empirical findings. The knowledge approach, for instance, nicely explains facts about cognitive development, whereas the exemplar theory can elegantly capture certain results from laboratory studies of category learning. Murphy concludes that a mixed theory is the most promising. If you look at a wide enough range of phenomena, it turns out that prototype representations, exemplars and knowledge structures all play a role in our mental life.

There is another way to conceive of a mixed theory, however. Murphy is explicit from the start that he expects no substantive differences across domains: we should learn and encode concepts such as 'tomato' in much the same way as concepts corresponding to events, linguistic categories, artistic styles, emotions, mathematical categories and so on. This is the standard assumption in the field, but it might be mistaken. Domains might matter. We know, after all, that there are some domain differences. If you consider geometrical concepts such as 'square', for example, the classical view is no longer so obviously a flop, whereas if you study abstract concepts such as 'mortgage' and 'democracy', no one could seriously doubt the role of higher-level knowledge.

In 1981, Smith and Medin divided their book by theories — their chapters had titles such as "The classical view", and "The exemplar view". In 2002, Murphy organized his book in terms of research issues, with chapters such as "Concepts in infancy" and "Conceptual combination". It is too early to tell, but perhaps the authoritative text in the 2020s will use a third way of dividing

up the field, with chapters such as "Natural kind concepts", "Social concepts" and "Mathematical concepts". ■

Paul Bloom is in the Department of Psychology, Yale University, New Haven, Connecticut 06520, USA.

The ethics of genetics

Genetics and Society: An Introduction

by Alison Pilnick

Open University Press: 2002. 224 pp.
£50, \$78 (hbk); £16.99 (pbk)

Genetic Politics: From Eugenics to Genome

by Anne Kerr & Tom Shakespeare

New Clarion Press: 2002. 224 pp.
£25, \$49.95 (hbk); £12.95, \$24.95 (pbk)

Bonnie Steinbock

Genetics always seems to be in the headlines, with the sequencing of the human genome, genetically modified foods, and the cloning of Dolly the sheep among the topics highlighted. The impact of these scientific developments on society is the subject of both of these books, but there the resemblance ends.

Genetics and Society by Alison Pilnick is a textbook and a reference material. It does not have a thesis, but rather presents scientific information and different viewpoints on a wide range of issues, including antenatal screening and testing, pharmacogenetics and genetically modified foods. This is a great deal to cover in just over 200 pages, and as a

result, many issues, such as intelligence (IQ) testing, receive a fairly superficial treatment.

Pilnick does a better job with the chapter on antenatal screening and testing, and a first-rate job in the chapter on genetically modified foods. There she evaluates, rather than simply presents, the opposing arguments, which makes for a more interesting chapter.

However, the book is marred by careless errors. For example, Pilnick claims that amniocentesis is twice as risky as chorionic villus sampling (the reverse is true). She mischaracterizes 'wrongful birth' suits as implicating the child's right not to be born, when it is the parents' right to terminate a pregnancy that has been abridged. Finally, she repeats the common misinterpretation of Immanuel Kant's principle "that an individual should never be thought of as a means, but always as an end". Kant's dictum does not refer to how we think of others, but to how we treat them. More importantly, it is not treating others as means that is prohibited, but treating them as mere means to our ends. It would be impossible never to treat other people as means, nor is this morally required by kantian ethics.

Genetic Politics by Anne Kerr and Tom Shakespeare does have a thesis: that the values and practices of eugenics continue to shape genetics today. The authors' approach to medical science in general and genetics in particular is "based on a commitment to radical critique of the cultural values and social relationships that shape knowledge and technology".

The disability-rights approach that they espouse is characterized by a rejection of the 'medical model'. Disability, according to this view, is not something to be prevented or cured by medical means. Instead, disability advocates opt for a 'socio-political approach' that regards the restrictions imposed by a disability as a function of societal arrangements, and not as something intrinsic to the condition. According to Kerr and Shakespeare: "There are serious moral and social problems with the drive to eliminate disease and difference at all costs. We do not believe that the world would be a better place without disabled people. Nor do we believe that disability is a tragedy best avoided."

If this claim is taken seriously, then it is not merely prenatal testing and selective abortion that are morally problematic, but also preventive measures, such as putting folic acid in flour to reduce the incidence of spina bifida. Indeed, if disability is not a "tragedy best avoided", it makes no sense to encourage pregnant women not to drink alcohol because of the risk of fetal alcohol syndrome (FAS), which results in mental retardation. Babies with FAS will be, as the authors characterize people with Down's syndrome, "just different".

The book usefully explains the complexity

Gone but not forgotten

The dodo was wiped out shortly after Dutch and Portuguese sailors arrived on Mauritius in the sixteenth century. But it wasn't until Lewis Carroll's classic *Alice's Adventures in Wonderland* that this strange-looking bird, shown here as it appeared in the frontispiece to Richard Owen's *Memoir on the Dodo* in 1866, gripped the public imagination. Erroll Fuller's *Dodo: From Extinction to Icon* (HarperCollins, £16.99) provides details of the dodo's natural history, behaviour and extinction, and tells how it became an icon for all things that have been lost forever.



of gene–gene and gene–environment interactions, which make the prediction of genetic diseases, much less behavioural traits, extremely complex. The authors provide convincing arguments against genetic reductionism. In addition, they point out the limits of justifying prenatal diagnosis as a matter of “choice”, when there are still so many constraints on choice. Certainly, women deciding whether to be tested for diseases in pregnancy, and whether to terminate, should be given objective, unbiased information

about living with a disability and raising a child with a disability. Their decisions should not be determined by a failure of society to provide the required resources. Nevertheless, even in an ideal society with bountiful resources, some couples and women will want to be tested and will opt for abortion to avoid having a child with a disability.

Kerr and Shakespeare argue that prenatal testing for disability leads to prejudice against, and a decrease in services for, disabled people. However, they provide no

empirical evidence for this claim, and indeed the reverse seems to be true. In the United States, disability-rights legislation, such as the Americans with Disabilities Act, has coincided with a rise in prenatal testing. There is no reason why society cannot prevent disability by a variety of means, including selective abortion, while also providing for the needs of those who are disabled. ■

Bonnie Steinbock is in the Department of Philosophy, University at Albany/State University of New York, Albany, New York 12222, USA.

Science in culture

A mechanical mind

Wolfgang von Kempelen's chess-playing automaton of 1770.

Martin Kemp

The possibility of creating a ‘human machine’, or more precisely a computer-driven device that exhibits independent ‘consciousness’, is one of the biggest remaining challenges for science. Although we tend to assume that the concept was first debated with the advent of computers — perhaps with a backward glance towards such pioneering devices as Charles Babbage’s ‘Analytical Engine’ — automata in fact have a history spanning centuries.

The terms of the debate had been framed as early as 1637, when René Descartes speculated in his *Discourse on Method* that: “If there were such machines with the organs and shape of monkeys or of some other non-rational animal, we would have no way of discovering that they were not the same as these animals. But if there were machines that resembled our bodies and if they imitated our actions as much as is morally possible, we would always have two very certain ways for recognizing that, nonetheless, they are not genuinely human. The first is that they would never be able to use speech... The second way is that ... one would discover that they did not act on the basis of knowledge, but merely as a result of the disposition of their organs.”

The making of astonishingly sophisticated automata in the eighteenth century brought Descartes’ model closer to realization. Most notable were Jacques de Vaucanson’s renowned devices, such as a duck that actively pecked grain, digesting it in a chemical stomach before egesting the results. They found their philosophical equivalent in Julien Offroy de La Mettrie’s scandalously godless tract *L’Homme-machine* in 1747, which undermined Descartes’ confidence in the separateness of humankind.

A practical answer to the objection that a machine could never be flexibly responsive seemed to be provided by the famed chess-playing automaton by Wolfgang von Kempelen, immortalized by Carl Gottlieb von Windisch in his 1784 book *Inanimate Reason; or A Circumstantial Account of that Astonishing Piece of Mechanism, M. de Kempelen's Chess Player*.

A Hungarian nobleman and prominent state official in Vienna, von Kempelen’s interests ranged from philosophy to large-scale engineering. In 1770 he stunned Maria Theresa’s court by unveiling his life-sized chess-player. Dressed in exotic Turkish garb, the robot sat behind a cabinet of machinery, which the presenter displayed by opening its doors.

The Turk — which perished in a fire in Philadelphia in 1854 — confounded spectators, defeating accomplished players, including Benjamin Franklin, and even becoming testy when opponents cheated. When he performed in Paris, one observer claimed that the device “is to the mind and eyes what M. Vaucanson’s flute player is to the ear”, a reference to the musical automaton that Vaucanson presented to the French academy in 1738.

The context went far beyond mere entertainment. In *The Mechanism of Speech in the Light of the Description of a Speaking Machine* (1791), von Kempelen made substantial claims about the communicative ability and consciousness of animals. If the animal-machine were capable of far more than the cartesians believed, why should a machine not play chess?

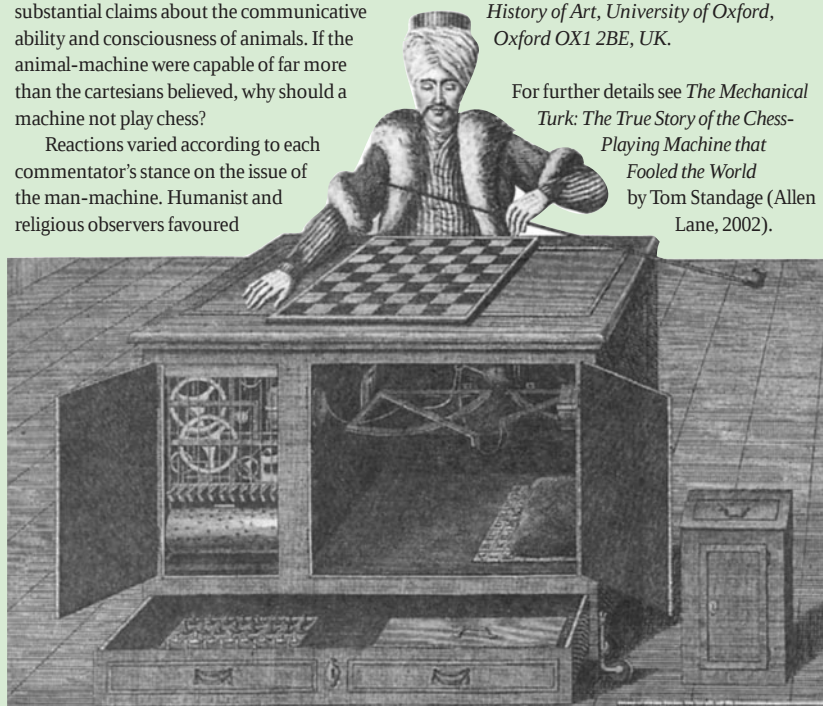
Reactions varied according to each commentator’s stance on the issue of the man-machine. Humanist and religious observers favoured

theories that a living player must somehow be concealed inside the cabinet. Those disposed to believe in the potential of machines and animals for performing elaborate mental tasks were prepared to credit the chess player with intelligence. Amongst those captivated by the machine’s implications were Babbage and Edgar Allan Poe.

How did it work? Johann Maelzel, who restored the Turk after von Kempelen’s death in 1804 and took it on tour, openly acknowledged that the Turk relied on a magician’s skills, akin to those at work in the illusion of the lady sawn in half. The secret lay in the cramped repositioning of a hidden player as the cabinet’s doors were opened in strict sequence. Although mechanically ingenious, it relied upon trickery. A chess-playing machine is, of course, now a formidable reality. It remains to be seen whether the bigger trick of a human machine can be performed for real.

Martin Kemp is in the Department of the History of Art, University of Oxford, Oxford OX1 2BE, UK.

For further details see *The Mechanical Turk: The True Story of the Chess-Playing Machine that Fooled the World* by Tom Standage (Allen Lane, 2002).



‘The Chess-Playing Turk’ from Carl Gottlieb von Windisch’s 1784 book *Inanimate Reason*.

All that jazz...

Nature's insight into what makes scientists tick.

Summarize yourself in the form of a title of a paper in Nature.

Slow explosion of jumping crocodiles enhanced by phosphorylation of the ARHGAP8 superconductivity complex. (Complex issues require complex titles.)

What was your first experiment as a child?

Building an airplane from our old washing machine. (It failed.)

Who have been the most important mentors in your career?

Michael Bölker and Svante Pääbo. Michael Bölker was my first direct contact in a lab. His fast and accurate analytical style of thinking taught me the basics in molecular biology and how to get from experiment to the understanding of a process. And without Svante Pääbo, I would not be in a position to be asked to write something about important mentors in my career. He has the courage and creativity to get into new fields before most people think of them.

Whose graduate student would you most like to have been (historical impossibility notwithstanding)?

Linus Pauling or Erwin Schrödinger.

What single scientific paper or talk changed your career path?

'Life with 6000 genes' (A. Goffeau *et al. Science* **274**, 546, 563–567; 1996). I remember reading this and realizing that 'genomic biology' is exciting and new, and that I definitely wanted to be a part of it.

What gives you the most job satisfaction now? What are your major frustrations?

Satisfaction lies in the understanding of a process, and the process of how one communicates this understanding. But when the process is over, frustration sets in. It is maybe like finding a song, working on it, playing it — and then having to play it over and over again.

What literary character would you employ as a postdoc?

Spider-Man.

You have the audience in your hands, but some smart-alec asks you the killer question you have no idea how to answer. What's your favourite response?

"This is a killer question and I have no idea how to answer it."

What book currently sits on your bedside table?

At the moment, *Intelligente Musik und wie*

man dazu Liebe macht (Intelligent music and how to make love to it) by Markus Wilmsmann. The author is my room-mate.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Goethe.

What music heads the playlist in your car or lab?

I generally listen to a lot of modern jazz. One of my favourite musicians is Bill Frisell, a jazz/avant-garde guitar player based in New York.

Where and when would you most liked to have lived or worked?

I'm very happy with here and now.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

Shut up, order a drink and listen.

What one thing would you rescue from your burning laboratory?

My dog.

What is the best piece of advice you've ever received?

'Don't listen to advice.' (I didn't listen to this advice.)

What do you most dislike about having research published?

The fixation in print of an ongoing process.

What would you have become, if not a scientist?

A jobless guitar player. When I was 21, I played electric guitar for about five or six hours a day (rock, blues, a little bit of jazz) and I thought I might turn professional. But I started too late, had not quite enough talent and later got fascinated by biology. So I decided to keep my freedom on the music side at the cost of spending less time on it.

Name one extravagance you can now get away with because of your eminence.

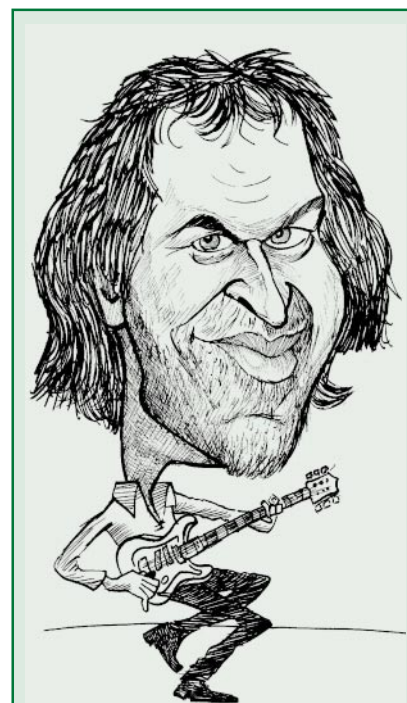
Not having to pick up the equipment from the rehearsal room before a concert.

What single discovery, invention or innovation would most improve your life?

A flying carpet.

What music would you have played at your funeral?

Platonic years by Nils Petter Molvaer from his



Wolfgang Enard

Rather than being a jobless guitar player, Wolfgang Enard works on molecular comparisons of humans and chimpanzees at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany.

album *Khmer*. It creates this state of happiness and sadness at the same time, which seems right for funerals.

Is there a 'tyranny of reductionism' in how scientists are trained today? That is, are students taught to look more into the workings of things and not enough at the 'big picture'?

Generally, I think it is right to train students so that they become experts in one field. However, in order to be able to communicate one's expertise it is crucial to place this field in the context of the 'big picture'. A wider appreciation of the history and philosophy of science might help in that direction.

What's the one thing about science that you wish the public understood better?

That natural science can tell you how things are but not how things should be.

If you could direct more government funding into one area of science (apart from your own, of course), where would you put it?

New approaches to gather data in order to compare the histories of different cultures.

What's just around the corner?

A kebab shop.

Branching out

Robert L. Charlebois,
Robert G. Beiko and Mark A. Ragan

Seven years after the publication of the first microbial genome sequence — that of *Haemophilus influenzae* — the roster of microbial genomes has topped 100. Despite early fears that whole-genome sequencing might be economically justified only for human pathogens, this list represents a gratifyingly broad range of microbial phenotypes — soil bacteria and photosynthesizers, thermophiles and halophiles, animal and plant pathogens, and more. At least 12 prokaryotic phyla are represented, as are a few eukaryotes — enough to allow a meaningful examination of the Tree of (microbial) Life.

In the early days of molecular phylogenetics (the mid-1960s to the early 1990s), it was thought that sequencing was the path to enlightenment — more sequences of more genes could only improve the depth and resolution of our knowledge of life's history. But instead, our 100-genome world is riven by seemingly irreconcilable conflicts; ambiguities and discrepancies are the norm, rather than the exception. Some of modern biology's fundamental tenets — notably the darwinian-mendelian model of parent-to-offspring ('vertical') gene flow — have once again, at least for microbes, been thrown into doubt. Lateral (horizontal) gene flow — in which genes are transmitted across, rather than along, branches in family trees — is no longer an explanation of last resort, but a competitive model for the origin of microbial biodiversity.

Although seldom correct, highly polarized views often serve to delineate a problem. Vertical inheritance with tree-like speciation fails to explain why so many gene families are distributed as they are among microbial genomes — that is, in highly diverse, sparse patterns that often fail to support accepted taxonomy. Yet genome evolution based largely on lateral (recombination-dependent) events seems

impossible for prokaryotes that live solitary lives inside eukaryotic cells, for example, or more generally reproduce for generation after generation by simple binary fission.

As recently as the 1970s, authoritative textbooks stated that microbiology might never be put on a phylogenetic footing. Carl Woese and colleagues showed, however, how the history of organismal life could be reconstructed from sequences of small-subunit ribosomal RNA, part of the protein-synthesizing machinery. The first molecular phylogenetic trees were sparse and not always trustworthy. But the 1980s brought automated technology for sequencing DNA, the polymerase chain reaction (PCR), and much-improved methods for inferring phylogenies. Ribosomal RNA (rRNA) genes are ubiquitous, with highly conserved termini that make amplification by PCR easy. Tens of thousands of rRNA sequences became available. Not surprisingly, the rRNA tree quickly became the 'gold standard' for determining microbial relationships.

Nonetheless, rRNAs provide only a narrow window on the microbial genome: for every gene that encodes an rRNA, there may be 1,000 that encode a protein. Protein-coding genes are less universal, more difficult to amplify by PCR, and often shorter and less information-rich than rRNA genes. What's more, trees inferred from individual protein-coding genes (or from their proteins) often disagree irreconcilably with the rRNA tree. Why is there such discrepancy, and what does it tell us about microbial genomes? These questions spark profound disagreement in the microbial-phylogenetics community.

Some theorists — let's call them the verticalists — remind us of the (real or supposed) inadequacies of single-gene phylogenetics. For verticalists, protein-based trees disagree because their true phylogenetic signal is too often obscured by noise and bias. Only by overcoming these obstacles — through using better models, perhaps, or cleaner data — can we understand how microbial genomes have diversified and evolved.

But others — the lateralists — point to the sophistication and power of existing methods, and argue that trees disagree because genes really do have different histories. Microbial genomes are, to a lateralist, more or less ephemeral entities that are maintained, if only fleetingly, by the vagaries of selection and chance. The apparent woesean hierarchy of taxa is only an epiphenomenon of differential barriers — whether environmental, geographical or more intrinsically biological — to lateral gene flow.

We and others have been exploring 'whole-genome trees' as a means of overcoming the noise and bias of single-protein

Microbial phylogenomics

Has genomics overturned the family tree of microbial life? Thanks in part to often polarized debate, elements of a new synthesis are emerging.

analyses, to extract the bulk phylogenetic signals that are inherent in genomes. The input data for genome trees can be the proportions of genes or proteins that genomes hold in common, or (as we prefer) the mean pairwise similarities between shared proteins. Despite some early indications to the contrary, whole-genome trees have now largely converged on the rRNA-sequence tree.

For us — as, presumably, for the verticalists — this convergence means that lateral gene transfer has not undermined descent with modification as the default explanation for microbial biodiversity, nor (as recently suggested by Ford Doolittle) has it thrown microbial classification into disarray. Lateral transfer is not both quantitatively important and directional. One of the few widely accepted instances of lateral gene transfer — the origin of chloroplasts from relatives of cyanobacteria — is clearly visible in our whole-genome trees, and even more so in 'sub-genome trees' based on functional subsets of genomes.

The most enthusiastic lateralists reply, however, that convergence between whole-genome and rRNA trees merely demonstrates that rRNA genes — unlike most individual protein-coding genes, but like the genome as a whole — are but pastiches that are produced by lateral gene transfer.

Fascinating as these conflicts are, the important point is not whether a given tree is right or wrong. Rather, we should use these trees as frameworks upon which to construct and test hypotheses about the rate and mode of microbial evolution, and to improve our analytical methods. Without conflicts, we might all be far more complacent about evolutionary theory. In microbial phylogenomics, the scientific process is alive and well! ■

Robert L. Charlebois and Robert G. Beiko are in the Department of Biology, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada. Mark A. Ragan is at the Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia.

FURTHER READING

- Clarke, G. D. P. *et al.* *J. Bacteriol.* **184**, 2072–2080 (2002).
Doolittle, W. F. *Science* **284**, 2124–2128 (1999).
Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. *Mol. Biol. Evol.* **19**, 2226–2238 (2002).
Woese, C. R. *Proc. Natl Acad. Sci. USA* **99**, 8742–8747 (2002).
Wolf, Y. I. *et al.* *BMC Evol. Biol.* **1**, 8 (2001).

P. MONDRIAN, THE GREY TREE/HANGS GEMEENTEMUSEUM/BRIDGEMAN ART LIBRARY



How close are we to a true impression of life's tree?

Sugars tied to the spot

Sabine L. Flitsch and Rein V. Ulijn

The interactions of sugars and proteins underlie many biological processes, and cataloguing them is a daunting task. A technique for attaching sugars to microarrays offers a promising, high-throughput solution.

Of the three main classes of biopolymers — proteins, nucleic acids and sugars — the sugars, or saccharides, are the most complex and hence the most difficult to study. Each monosaccharide building-block has multiple attachment sites, which means that sugars can be built in a variety of linear or branched fashions. For example, two common sugars, glucose and mannose, can be linked to form a disaccharide in up to 80 different ways. It takes just a few coupling steps to produce a large number of diverse biological structures, and this diversity of saccharide structure is exploited *in vivo*. A major challenge in cell biology is to decipher the sugar code, or 'glycome' — that is, to define the interactions between cell-coating sugars and proteins and work out how they recognize each other. Microarray techniques offer the possibility of high-throughput analysis and have been investigated in a series of reports^{1–3}, the latest by Fabio Fazio and colleagues in the *Journal of the American Chemical Society*⁴.

One of the best-known examples of medically important sugar–protein interactions is the action of the human blood-group antigens A, B and O. The difference between these three antigens lies in a single carbohydrate unit in an oligosaccharide sequence on the surface of red blood cells. When different blood types are mixed, any 'foreign' A, B or O antigen is recognized by highly specific sugar-binding proteins (agglutinins) that cause blood agglutination, or clotting. Hence, transfusions fail if the blood type is not matched to that of the patient. Other types of sugar-recognizing proteins include those activated to adhere to sugars in immune responses, lectins that crosslink cell-surface sugars, antibodies that are specific to certain sugars, and enzymes such as glycosidases and glycosyltransferases.

To investigate the great diversity of oligosaccharide structure requires an automated technique, but it should also be sensitive to small quantities of material. Microarrays have been used successfully in the study of gene expression and protein interactions, and are particularly well suited to studying the glycome. By attaching complex sugars to a microarray plate, the compounds can be quickly and easily screened to work out their specific interactions with proteins. But the problem is how to link the sugar under study to the microarray surface.

A method devised by Fukui *et al.*¹ relies on

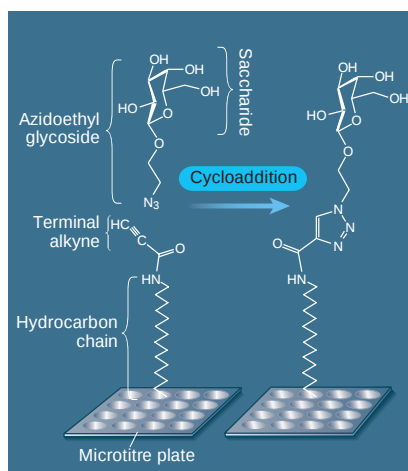


Figure 1 Making a microarray of sugars. Fazio *et al.*⁴ exploit a simple cycloaddition reaction to attach complex sugars to a microtitre plate. The alkyne group at the end of a hydrocarbon chain (attached to the plate) is a natural tether for azidoethyl glycoside, whose terminating N₃ group binds to the alkyne, forming a ring.

the fact that the reducing end of all oligosaccharides contains a unique functional group (an aldehyde) that reacts with an aminolipid to form a so-called neoglycolipid by 'covalent conjugation'. The neoglycolipids can then be immobilized efficiently by spotting them into arrays on nitrocellulose

membranes. In an ambitious experiment, Fukui *et al.* arrayed 30 different fractions of neoglycolipids generated from rabbit-brain polysaccharides (O-linked glycans) on nitrocellulose. Using highly specific antibodies to probe for certain carbohydrate sequences, they were able to obtain structural information on the oligosaccharides present.

Wang and colleagues² have shown that carbohydrates can also be spotted directly on single nitrocellulose-coated glass slides without the need for covalent conjugation. In this system, up to 20,000 spots of microbial antigens could be printed on a slide — which would be enough to include most of the known human microbial antigens. The sensitivity is such that antibodies can be detected if the microarray is tested with only one microlitre of human blood serum.

But there are drawbacks to each of these techniques: in the neoglycolipid approach, the first monosaccharide unit is lost on binding to the nitrocellulose membrane, and the Wang *et al.* method gives little control over the orientation of the immobilized polysaccharide. An exciting new development is the use of cycloaddition reactions — a class of simple, selective chemical reactions in which a ring structure is formed — by Houseman and Mrksich³, and now by Fazio *et al.*⁴. Using a specific cycloaddition, the Diels–Alder reaction, Houseman and Mrksich showed that a series of ten different

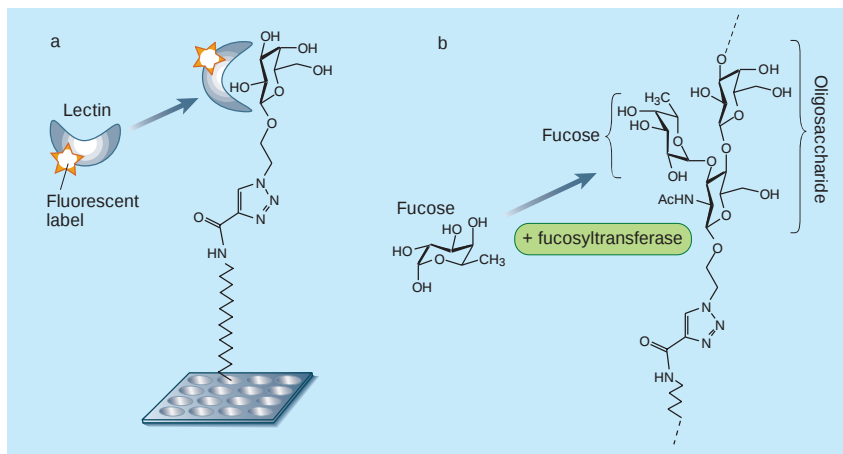


Figure 2 Putting the sugar-microarray technique⁴ to the test. Fazio *et al.* provide several examples to demonstrate the potential of their approach for high-throughput analysis of protein–sugar interactions. These include: a, the binding of lectin molecules to sugars on the microtitre plate traced through fluorescein labelling; and b, addition of a fucose group, mediated by the enzyme fucosyltransferase (and monitored using a fucose-specific lectin, not shown).

monosaccharides could be specifically detected by five different lectins. Fazio *et al.* exploit cycloaddition between azides (which contain an N₃ group) and alkynes.

Fazio *et al.* made microarrays of oligosaccharides by immobilizing the molecules on polystyrene microtitre plates. Starting with azidoethyl glycosides, they exploited the cycloaddition reaction with the terminal alkyne group of a saturated hydrocarbon chain to bind the sugars to the microtitre plate (Fig. 1). In a series of experiments, 11 different di-, tri- and tetrasaccharides were successfully attached to the plates. Fazio *et al.* then tested their behaviour as ligands for lectins (Fig. 2a), tracking the binding of these proteins by labelling them with the fluorescent marker fluorescein (C₂₀H₁₂O₅).

The authors also demonstrated that the oligosaccharides attached to the microarray were still accessible as substrates for enzymatic transformations such as fucosylation, in which a fucose group is attached to the oligosaccharide through the action of fucosyltransferase (Fig. 2b). Fucosylation of a particular oligosaccharide produces the tetrasaccharide 'sialyl Lewis x', which is known to be involved in inflammation: this tetrasaccharide is produced on leukocytes (white blood cells) as a response to inflammatory stimuli and then mediates, through protein–sugar interactions, the recruitment of leukocytes from the bloodstream to

inflammatory lesions. Fazio *et al.* tracked the process of fucosylation on their microarray using a fucose-specific lectin, and confirmed the results by mass spectrometry.

This work shows the potential for sugar-microarray technology, which in the future should permit systematic and high-throughput analyses of protein–sugar interactions and enzyme-inhibition screening, similar to the analyses being developed for nucleic acids and proteins. At the moment, there is still the problem of chemically generating diverse libraries of saccharides to work on, but efforts to automate oligosaccharide synthesis are under way. With microarrays, it should ultimately be possible to compare sugar expression on different proteins, cells or organs, and even to monitor how the sugar patterns on cell surfaces change as the cells develop and take on their particular function in the living system. ■

Sabine L. Flitsch and Rein V. Uljijn are in the School of Chemistry, University of Edinburgh, Joseph Black Building, West Mains Road, Edinburgh EH9 3JJ, UK.
e-mails: sabine.flitsch@ed.ac.uk
rein.uljijn@ed.ac.uk

1. Fukui, S., Feizi, T., Galustian, C., Lawson, A. M. & Chai, W. *Nature Biotechnol.* **20**, 1011–1017 (2002).
2. Wang, D., Liu, S., Trummer, B. J., Deng, C. & Weng, A. *Nature Biotechnol.* **20**, 275–281 (2002).
3. Houseman, B. T. & Mirksich, M. *Chem. Biol.* **9**, 443–454 (2002).
4. Fazio, F., Bryan, M. C., Blixt, O., Paulson, J. C. & Wong, C.-H. *J. Am. Chem. Soc.* **124**, 14397–14402 (2002).

Functional genomics

RNA sets the standard

Thomas Tuschl

One way of finding out what genes do is to inactivate them, and to study the effects, in 'model' organisms. That has now been done for many thousands of worm genes in two large-scale analyses.

Genome-sequencing projects provide a tremendous amount of information about the genetic make-up of an organism. One of the most obvious types of data to be gleaned from these ventures is a list of genes: for instance, the nematode worm *Caenorhabditis elegans* — the first animal to have its genome sequenced — has a predicted 19,757 genes that code for proteins. A list won't tell you what these proteins do, however, and one way of starting to find out is to inactivate the genes, one at a time, and see what happens. As they describe on page 231 of this issue, Kamath and colleagues¹ have done just that, using double-stranded RNAs to rapidly and transiently inactivate 16,757 of the worm's predicted protein-coding genes. Meanwhile, Ashrafi and co-workers² (page 268) have analysed these genes specifically to see if they have a role in regulating body fat. Together, their work sets a new standard for systematic, genome-wide genetic studies.

The analysis of gene function in animals and plants was revolutionized in 1998 by the discovery³ of the mechanism underlying 'RNA-mediated interference' (RNAi) in *C. elegans*. In its natural form, this is a cellular defence process that protects against genome-invading transposable genetic elements and viruses. RNAi is now widely used by researchers to temporarily block gene expression; its effects on the morphology, physiology and behaviour — the phenotype — of an organism are similar to those of classical gene knockouts (deletions of segments of a gene's protein-coding regions)⁴.

In nature, the trigger for this cellular defence mechanism is double-stranded RNA (dsRNA), which is normally produced from uncontrollably replicating invasive genetic elements but never from tightly regulated cellular genes. The nucleotide sequence of the dsRNA guides the defence machinery towards single-stranded messenger RNAs of

the same sequence as the dsRNA, and directs their destruction. The generation of an mRNA is the first step in producing a protein from a gene, so, in the case of invasive genetic elements, destroying mRNA abolishes the synthesis of crucial proteins, such as those needed for replication. Researchers have taken advantage of this basic mechanism: introducing a dsRNA identical in sequence to a cellular gene of choice will cause the destruction of the cellular mRNA, thus disrupting protein synthesis.

Caenorhabditis elegans is remarkable in that such 'gene silencing' can result when the dsRNA has simply been eaten⁵. The bacterium *Escherichia coli* is the preferred diet of *C. elegans*, and molecular-genetic tools allow appropriate dsRNA-encoding DNA sequences (plasmids) to be introduced into *E. coli* (Fig. 1). The effects of the loss of function of a gene are best observed in the offspring (either embryos or larvae) of hermaphrodite worms that have feasted on bacteria modified in this way. In the past, this and similar approaches have been used to target about one-third of all *C. elegans* genes^{6–8}. Kamath *et al.*¹ have now almost tripled the amount of available information.

The authors show that inactivation of about 10% of the genes they studied produced a defective phenotype: the offspring might die as embryos or larvae; or the larvae might grow slowly, be sterile, have morphological defects, or move in an uncoordinated way. One interesting finding was that targeting *C. elegans* genes for which there are conserved counterparts in other eukaryotes (such as yeast, fruitflies and the plant *Arabidopsis thaliana*) was more likely to produce an aberrant phenotype than targeting less conserved genes. That might have been expected, as genes that are conserved during evolution are likely to be essential. Also, a peculiarity of the *C. elegans* genome is that it contains many recently duplicated genes. Kamath *et al.* show that gene duplicates are at least partially functionally redundant, or have such specialized functions that defective phenotypes are not detectable. Simultaneous inactivation of such sequence-related genes may be necessary to uncover their functions.

The authors also analysed the protein structural domains encoded by the genes they studied. Genes that tended to have non-lethal effects when inactivated by RNAi generally encoded domains that were 'animal-specific', being found in either humans or fruitflies as well as *C. elegans*. By contrast, genes that caused death when inactivated generally encoded domains that are of more ancient origin, being found in yeast and *Arabidopsis* as well. So, the domains that are essential for survival have been preserved across evolutionarily diverse species; the domains that are not needed for survival appear to have evolved during

animal evolution to produce animal-specific functions, such as movement.

Another intriguing finding was that inactivating most of the genes on the X chromosome had no effect on embryonic or larval survival, although many of these genes (as many as on non-sex chromosome V) did produce defective phenotypes when silenced. The authors propose that this is because many X-chromosome genes are naturally switched off during development, being turned on only later. Finally, Kamath *et al.* show that the genes that produce a noticeably defective phenotype when inactivated tend to be clustered in large chromosomal regions of up to 1 million base pairs. That suggests some form of long-range coordinated regulation of these genes.

The libraries of plasmids and modified *E. coli* developed by Kamath *et al.*¹ provide a powerful tool for loss-of-function genetic screening, especially when geared towards a particular molecular or phenotypic read-out. In the second paper (which has many of the same authors), Ashrafi *et al.*² describe such a screen for fat-regulatory genes in *C. elegans*. As well as the dsRNA, the authors included a fluorescent dye in the worms' diet. This allows fat droplets in intestinal cells of living worms to be visualized. They then determined the amount of body fat by measuring the fluorescence intensity.

The authors identified 305 genes that reduced the amount of body fat upon inactivation, and 112 genes that increased it. By repeating the RNAi screen for these genes in worms whose body fat was already altered by mutation, Ashrafi *et al.* were able to assign some of the genes to known fat-regulating pathways. Many of these genes are conserved in humans, so might represent new targets for anti-obesity drugs. Other functional RNAi-based screens are to be anticipated, and the worm could become a major model organism for identifying new candidate drugs to treat metabolic diseases.

RNAi-based loss-of-function screens like these^{1,2} are tremendously powerful. Yet they have some disadvantages compared with classical genetic screening. For instance, classical screening also looks at what happens when a gene's activity is turned up; such studies have often proved essential for ordering genes in pathways. Also, proteins are generally embedded in a network of interactions with many partners, and more highly specific phenotypes are expected if only one of the protein's interaction domains is altered by mutation, instead of the entire molecule being eliminated by RNAi. Moreover, some genes are more difficult to target by RNAi than others. And there are many non-coding RNAs, which are not translated into proteins; it remains to be seen if they are susceptible to RNAi.

But there is still a vast number of protein-coding genes, from many different organisms,

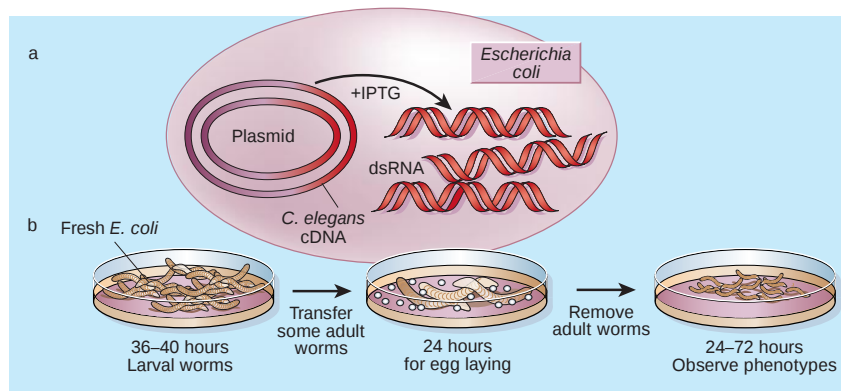


Figure 1 Gene screening by double-stranded-RNA-mediated interference (RNAi). Kamath *et al.*¹ and Ashrafi *et al.*² used the following technique to silence the expression of 16,757 genes individually in *Caenorhabditis elegans*. a, DNA molecules (plasmids) encoding a double-stranded RNA (dsRNA) of choice are inserted into *Escherichia coli* bacteria. Incubation with isopropylthio- β -galactoside (IPTG) induces production of the dsRNA. b, Worms at the latest larval stage are placed on a lawn of *E. coli*, and allowed to feed. Several adult worms are then placed onto new plates seeded with the same bacteria to lay eggs. The offspring are monitored for embryonic death and post-embryonic phenotypes, such as slow larval growth or movement disorders.

to study in detailed RNAi-based functional analyses, and this will keep the army of cell and molecular biologists busy for some time. Soon it will be possible to carry out RNAi-based screens in animal and human cells, using short synthetic double-stranded RNAs³ or plasmid- or virus-based DNA molecules that encode hairpin RNAs¹⁰. With the development of phenotypic read-outs based on cell biology, the hunt will begin.

Thomas Tuschl is in the Laboratory of RNA Molecular Biology, Rockefeller University,

1230 York Avenue, New York, New York 10021, USA.
e-mail: ttuschl@mail.rockefeller.edu

1. Kamath, R. S. *et al.* *Nature* **421**, 231–237 (2003).
2. Ashrafi, K. *et al.* *Nature* **421**, 268–272 (2003).
3. Fire, A. *et al.* *Nature* **391**, 806–811 (1998).
4. Plasterk, R. H. *Science* **296**, 1263–1265 (2002).
5. Timmons, L. & Fire, A. *Nature* **395**, 854 (1998).
6. Fraser, A. G. *et al.* *Nature* **408**, 325–330 (2000).
7. Gönczy, P. *et al.* *Nature* **408**, 331–336 (2000).
8. Maeda, I., Kohara, Y., Yamamoto, M. & Sugimoto, A. *Curr. Biol.* **11**, 171–176 (2001).
9. McManus, M. T. & Sharp, P. A. *Nature Rev. Genet.* **3**, 737–747 (2002).
10. Tuschl, T. *Nature Biotechnol.* **20**, 446–448 (2002).

Palaeoclimatology

Cooling a continent

Peter Barrett

The effect of greenhouse gases on climate is underscored by modelling work showing that formation of the Antarctic ice sheet, 34 million years ago, occurred largely because of a fall in atmospheric CO₂ concentration.

The first continent-wide Antarctic ice sheet formed in earliest Oligocene times¹, now dated at about 34 million years ago. Given its global significance, and that the past is one of the few ways we have of peering into our climatic future, understanding that event has been among the main aims in palaeoclimatology. On page 245 of this issue, DeConto and Pollard² provide an account that differs from the commonly accepted explanation in invoking changes in atmospheric CO₂, rather than in ocean circulation, as the determining factor.

The previous scientific story goes back to the 1970s, when Shackleton and Kennett³ pioneered the recovery of long-term temperature data from deep-sea sediments, in the form of oxygen isotope ratios preserved in calcium carbonate microfossils. They identi-

fied a progressive global cooling over the past 65 million years, with step-like falls in temperature at about 34 and 15 million years ago, as shown in Fig. 1, overleaf⁴. Kennett⁵ linked these steps to the separation of Australia and South America from Antarctica, which opened ocean gateways that forced current systems in the Southern Hemisphere from primarily north–south flows in the Pacific, Atlantic and Indian Oceans to west–east flows around Antarctica. The inferred consequence was cooling of the Antarctic region by thermal isolation, through a reduction in heat transfer between the Equator and the South Pole, permitting sea ice to form. Subsequent further cooling allowed the formation of a large permanent ice sheet, although sediment cores from the Antarctic margin have since shown that the first cooling

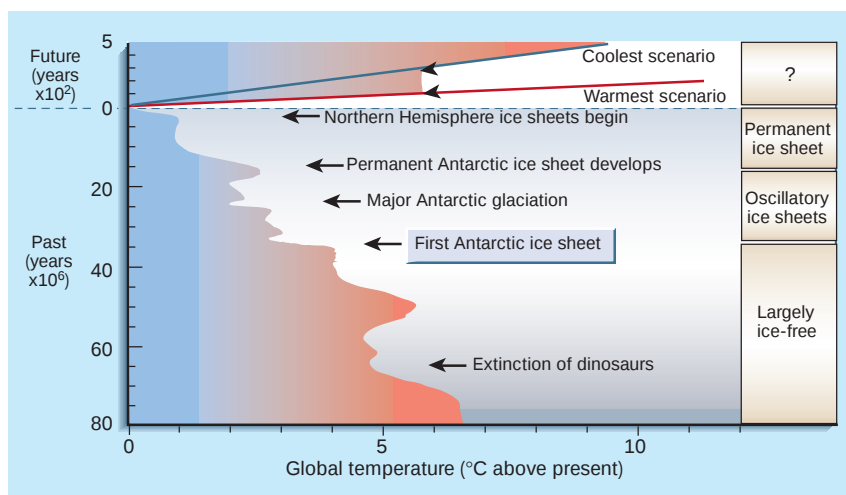


Figure 1 Decline in average global temperature over the past 80 million years of Earth's history. This global temperature curve is inferred from deep-sea isotope data⁴, the zero point being today's temperature of 15 °C. Of the significant transitions in glacial history⁷ shown, the sharp cooling associated with the formation of the first Antarctic ice sheet, 34 million years ago, is ascribed by DeConto and Pollard² to a decrease in the concentration of atmospheric CO₂. Global temperature projections for the future, produced by the Intergovernmental Panel on Climate Change (IPCC)¹⁴, are shown on a greatly expanded timescale. According to the warmest IPCC scenario, by the end of this century global climate will be similar to that near the end of the reign of the dinosaurs. On the coolest scenario, this climate will be achieved in about 300 years, assuming that the CO₂ concentration continues to increase at the present rate. The glacial oscillations that have occurred in the past on timescales of 40,000 and 100,000 years¹ are not shown, but each decreased the global temperature by about 5 °C.

step, 34 million years ago, was sufficient for extensive ice growth on the continent¹.

DeConto and Pollard² report modelling studies of two further influences on the Antarctic climate—changes in atmospheric CO₂ levels and in the Earth's orbital parameters. They show how a decline in CO₂ content, from four times to double that of our atmosphere before the past century (a baseline termed the pre-industrial atmospheric level), can result in a climate in which winter snow eventually survives through the summer and over a large enough area to start forming an ice sheet. They also show how the exact timing of ice-sheet initiation can be affected by a combination of the Earth's orbital parameters that results in minimal summer insolation of the Antarctic.

DeConto and Pollard also checked on the likely influence, previously queried⁶, of the ocean gateways opening south of Australia and South America at about the time that the first big ice sheet grew. To do this they compared the effects on simulated ice-sheet growth of an open and a closed Drake Passage, using the estimate of 20% for the reduction in southward ocean heat transport resulting from gateway opening. The Drake Passage lies between South America and the Antarctic Peninsula, and when closed represented the final barrier to circum-Antarctic circulation. The result of the experiment showed a shift in the threshold CO₂ level for ice-sheet growth from 2.4 (closed) to 2.8 (open) times pre-industrial levels. However,

there was no significant change in the pattern or magnitude of ice-sheet formation, confirming the primacy of atmospheric CO₂ as the controlling factor.

This study² is a fine example of how modern computer technology can be applied to major questions about climate processes and history. It also highlights areas where further research is needed, such as the relative influence of temperature and ice volume on the deep-sea isotope curve, and the effects of the roughness and hardness of the terrain beneath an ice sheet on its behaviour. But as with all models of past events, however elegant, it requires validation through concrete evidence from observations of today's world, and the records in ice and sediment layers on land and beneath the sea. For the first Antarctic ice sheet, the relevant timing and geography have been fairly well established from the evidence cited by DeConto and Pollard, and the orbital variations they call upon to trigger the first ice sheet are recorded in younger, cyclic sequences evident in cores drilled off the Antarctic margin⁷. However, a sedimentary record of the transition from 'greenhouse' to 'icehouse' Antarctica on the continent itself, with information on the extent and rate of the change in climate and vegetation, has yet to be recovered. Drilling programmes such as ANDRILL⁸ and SHALDRIL⁹, now under way, should provide the relevant data—for example on changes in coastal temperatures through the transition period.

What of other climate transitions in the

past? For instance, the small step at about 24 million years ago that coincides with a widely recognized fall in sea level¹⁰ and a shift in Antarctic coastal vegetation from beech forest to tundra¹¹? Or the larger step at 15 million years ago, which is widely interpreted as the transition from an ephemeral to a permanent Antarctic ice sheet¹²? These two episodes are still not understood. There is also continuing curiosity over the response of the Antarctic ice sheet during the worldwide warming about 3 million years ago. Here uncertainties need to be resolved in dating glacial deposits in the Transantarctic Mountains. On the basis of rare microfossils, some contend that the deposits date to this time¹²; others, however, have shown how the microfossils could be windborne contaminants, and the deposits could be much older¹³.

As to the future, it is now clear that land-surface and ocean temperatures are rising in response to human-induced emissions of greenhouse gases—and remarkably fast on a geological timescale. The present concentration of greenhouse gases is 30% above the benchmark pre-industrial content, rising to that level in only a century, and is likely to be the highest for the past 20 million years¹⁴. According to the 'warmest scenario' projections of the Intergovernmental Panel on Climate Change, the projected doubling of CO₂ concentration by the end of this century is expected to result in a climate that the planet last experienced about 70 million years ago (Fig. 1).

The effects of this change are difficult to predict, but they will plainly be profound. Finding out more will require closer collaboration between geoscientists who work at extracting data on past climates from ice and sediment cores and from rock strata, and modellers who use complex mathematical representations to approximate events in the physical world. An example of such collaboration is the formation of the Antarctic Climate Evolution planning group¹⁵.

More generally, DeConto and Pollard's study² brings a new understanding of the effects of CO₂ emissions on climate, and adds force to the arguments for reducing greenhouse-gas emissions beyond those agreed in the Kyoto Protocol. At the same time it is clear that regional responses to climate change through interactions between the oceans, atmosphere and ice sheets vary considerably from the global average. The science of mitigating the effects of greenhouse emissions is in some countries and cases taking priority over research into understanding climate behaviour, especially regionally. Both areas of research will need sustained support, along with an international commitment to an effective solution, if we are to survive the worst consequences of this grandest of all human experiments. ■

Peter Barrett is at the Antarctic Research Centre and School of Earth Sciences, Victoria University of

Wellington, Kelburn Parade, Wellington 6005,
New Zealand.

e-mail: peter.barrett@vuw.ac.nz

- Barrett, P. J. *Terra Antarctica* **3**, 103–119 (1996).
- DeConato, R. M. & Pollard, D. *Nature* **421**, 245–249 (2003).
- Shackleton, N. J. & Kennett, J. P. *Init. Rep. Deep Sea Drilling Project* **29**, 743–755 (1975).
- Crowley, T. J. & Kim, K.-Y. *Geophys. Res. Lett.* **22**, 933–936 (1995).
- Kennett, J. P. *J. Geophys. Res.* **82**, 3843–3859 (1977).
- Sloan, L. C., Walker, J. C. G. & Moore, T. C. Jr *Paleoceanography* **10**, 347–356 (1995).
- Naish, T. R. *et al.* *Nature* **413**, 719–723 (2001).
- <http://andril-server.unl.edu/index.htm>
- www.arf.fsu.edu/arhtml/arfpages/shaldril.html
- Miller, K. G., Wright, J. D. & Fairbanks, R. G. *J. Geophys. Res.* **96**, 6829–6848 (1991).
- Raine, J. I. & Askin, R. A. *Terra Antarctica* **8**, 389–400 (2001).
- Wilson, G. S. *et al.* *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **179**, 1–15 (2002).
- Barrett, P. J., Bleakley, N. L., Dickinson, W. W., Hannah, M. H. & Harper, M. A. in *The Antarctic Region: Geological Evolution and Processes* (ed. Ricci, C. A.) 763–770 (Terrapub, Siena, 1997).
- Houghton, J. T. *et al.* *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, 2001).
- www.geo.umass.edu/ace

Immunology

Mobilizing the army

Steven D. Shapiro

When our bodies are injured or infected, inflammatory cells migrate to the damaged area to carry out rescue and repair work. Interactions between three types of protein may form the basis of a highway to guide these cells.

Our bodies respond to injury and infection by mobilizing inflammatory cells, which, on reaching an afflicted area, kill microorganisms, eliminate any debris and regulate tissue repair. Neutrophils, which are generated in the bone marrow and circulate through the bloodstream, are the first such cells to arrive on the scene. Armed with potent protein-digesting enzymes and oxidants, neutrophils are consummate microbe killers. But if these cells accumulate and are activated in an uncontrolled way, they can cause excessive inflammation and injure the very tissues they are designed to protect. So it is important to understand how neutrophils find their way to, and so accumulate at, sites of injury. A breakthrough came recently with a paper published in *Cell* by Li and colleagues¹.

As in previous studies of tissue damage, Li *et al.*¹ infused the drug bleomycin into the lungs of mice; this causes short-lived injury and consequent inflammation of the layer of epithelial cells that lines the lungs, followed by long-lasting scarring. The authors show that when wild-type mice are treated in this way, neutrophils migrate through the lungs' capillaries, cross both the capillary wall and the epithelial layer, and enter the airspace, where they start to repair the damage. But the situation is rather different in mice lacking a particular matrix metalloproteinase² — an enzyme that cleaves components of the web of proteins and other molecules (the extracellular matrix) in which cells are embedded in tissues. Li *et al.* show that without this enzyme, called matrilysin or MMP-7, neutrophils accumulate in the interstitial space that separates capillaries from the epithelial tissue, and do not cross the epithelium or enter the airspace.

So MMP-7 is clearly necessary for neutrophils to migrate to sites of injury, at least in this case. But what is the key substrate for this enzyme? The authors went on to investigate,

and in so doing uncovered a mechanism by which three components of the epithelial tissue interact to produce a chemical (chemotactic) gradient that attracts migrating neutrophils (Fig. 1, overleaf). When injured, epithelial cells secrete a chemokine called KC; chemokines are proteins known for their ability to contribute to chemotactic gradients, attracting cells in the direction of the gradient. KC binds to syndecan-1, an adhesive component of the extracellular matrix that contains heparan sulphate. MMP-7 is also released from epithelial cells, and cleaves or 'sheds' the syndecan-1–KC complex from the extracellular matrix, allowing a gradient of this complex to form.

All of these protein families — matrix metalloproteinases, chemokines and matrix components — had already been individually implicated in cell movement. The triumph of Li *et al.*¹ is to explain how these molecules fit together *in vivo*, and to clarify their exact functions in neutrophil migration. For instance, the ability of many cells to move and to traverse tissue barriers has long been ascribed largely to matrix metalloproteinases, with tumour cells in particular using these enzymes to clear paths, promoting local tissue invasion and distant metastasis. Extrapolating from these observations of tumour cells, it was speculated that inflammatory cells use their cadre of matrix metalloproteinases to move through tissue barriers in a similar fashion. Indeed, this has been seen to occur.

Yet confusion remained, because it was obvious that these enzymes do more than simply clear road-blocks from the path of migrating neutrophils. Moreover, cells have other ways of moving. For example, tumour cells normally take straight paths, degrading the extracellular matrix as they go. But in the presence of matrix-metalloproteinase inhibitors, such cells change both their pattern of movement and their shape as they

pick their way through microscopic gaps in the matrix (P. Freidl, personal communication). Neutrophils, meanwhile, are known to be the most 'flexible' inflammatory cells, able to weave through tissue barriers in the absence of protein-digesting enzymes³. All in all, then, Li *et al.*'s discovery¹ that MMP-7 is involved in neutrophil migration was perhaps not surprising. But its precise role — in 'shedding' syndecan-1–KC complexes to form a chemical gradient — was unanticipated, and shows that MMP-7 is needed for directed neutrophil migration, rather than just movement *per se*.

To complicate the picture still further, matrix metalloproteinases may also limit inflammation. Three years ago, McQuibban *et al.*⁴ showed that some of these enzymes process and inactivate the chemokines to which other inflammatory cells — monocytes — respond, thereby restricting monocyte migration into sites of injury. This, together with the work of Li *et al.*, suggests a feedback mechanism whereby matrix metalloproteinases initially promote cell influx to damaged areas; later, excessive quantities of these enzymes turn off the pro-inflammatory chemokine signals.

And what of the chemokines? These proteins have long been known to act as attractants by binding to receptors on the surface of inflammatory and immune cells. But the large number of chemokines, and the lack of specificity of particular receptors for particular chemokines, has made it difficult to ascribe functions to individual molecules. So it is another significant feature of the new work¹ that, although the authors found many chemokines to be present in injured mouse lungs, a gradient of chemokine KC alone (established by syndecan-1) was both necessary and sufficient to control neutrophil migration.

Which brings us, finally, to syndecan-1. Fragments of other matrix proteins have previously been shown to be chemoattractants in cell cultures. For instance, fragments of elastin attract monocytes and fibroblast cells^{5,6}, whereas fragments of laminin, collagen or entactin attract neutrophils. These fragments are believed to work alone, but their significance *in vivo* is unclear. The importance of fragments of syndecan-1, however, is now apparent. Previously, it was shown that interleukin-8 (which is related to chemokine KC) must bind heparan sulphate to mediate directed neutrophil migration⁷, and that metalloproteinases can shed the extracellular domain of syndecans⁸. Li *et al.* tie all this work together, showing that syndecan-1–KC complexes are snipped away from the main extracellular matrix by matrix metalloproteinases. The broader implication is that other matrix fragments may cooperate with other chemokines to create gradients that direct cell movement.

So Li *et al.*'s work completes one part of

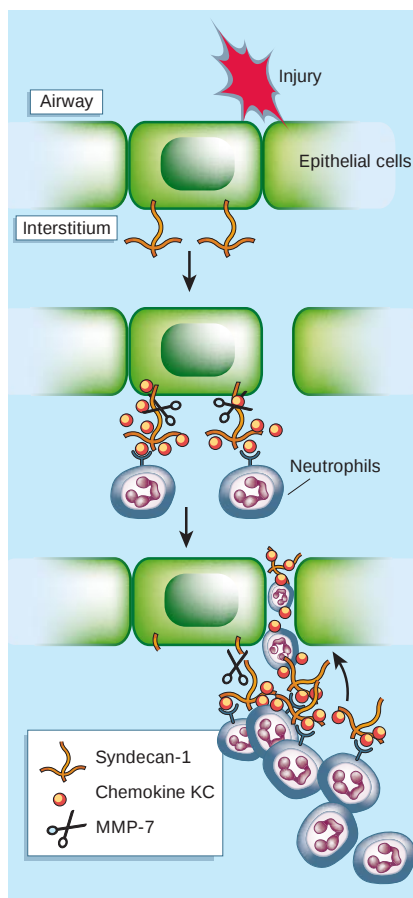


Figure 1 Directing neutrophils to sites of injury. When damaged, epithelial cells (such as those lining the lungs) secrete the chemokine KC, which binds to syndecan-1 on an extracellular matrix scaffold¹. Neutrophils then bind to KC. The epithelial cells also release MMP-7 (also known as matrilysin), which cleaves off the syndecan-1-KC complex. This forms a chemical gradient that directs neutrophils to the site of injury. (Figure adapted from one supplied by W. C. Parks.)

this jigsaw puzzle. But there are still several pieces to slot into place. Why, for instance, are so many chemokines and proteinases involved in moving a limited number of cell types? How are the production and interaction of these proteins controlled? And how exactly does the sticky syndecan-1, attached to a chemokine, guide cells to their destination following cleavage? The answers might allow us to control the inflammatory process, to improve the removal of microorganisms and the repair of tissues, while limiting damage. ■

Steven D. Shapiro is at the Brigham and Women's Hospital, Harvard Medical School, 15 Francis Street, Boston, Massachusetts 02115, USA.
e-mail: sshapiro@rics.bwh.harvard.edu

- Li, Q., Park, P. W., Wilson, C. L. & Parks, W. C. *Cell* **111**, 635–646 (2002).
- Wilson, C. L., Heppner, K. J., Labosky, P. A., Hogan, B. L. M. & Matrisian, L. M. *Proc. Natl Acad. Sci. USA* **94**, 1402–1407 (1997).

- Huber, A. R. & Weiss, S. J. *J. Clin. Invest.* **83**, 1122–1136 (1989).
- McQuibban, G. A. *et al. Science* **289**, 1202–1206 (2000).
- Senior, R. M., Griffin, G. L. & Mecham, R. P. *J. Clin. Invest.* **66**, 859–862 (1980).
- Hunninghake, G. W. *et al. Science* **212**, 925–992 (1981).

- Webb, L. M. C., Ehrenguber, M. U., Clark-Lewis, I., Baggiolini, M. & Rot, A. *Proc. Natl Acad. Sci. USA* **94**, 7158–7162 (1993).
- Fitzgerald, M. L., Wang, Z., Park, P. W., Murphy, G. & Bernfield, M. *J. Cell Biol.* **148**, 811–824 (2000).

Quantum cryptography

Code-breakers confounded

Mark Hillery

Coherent-state quantum cryptography holds the promise of efficient, secure communication. An experimental demonstration shows that a secure key to the code can be exchanged, even if there is a large transmission loss.

Quantum cryptography makes use of the unusual properties of quantum mechanics to protect encoded information. In trying to listen in on a message sent through a properly designed quantum communication channel, an eavesdropper will inevitably disturb the signal and thereby reveal his or her presence. The quantum cryptographic schemes that have been explored so far have made use of either single photons or very weak light pulses. Could it be possible to use more intense light pulses containing many photons and still take advantage of quantum mechanics to protect secret information? In their paper on page 238 of this issue, Frédéric Grosshans and colleagues¹ show that it is. They have constructed a table-top system that encodes information in particular quantum states of light that contain several hundred photons each, then transmits them and decodes the information. This could lead to a faster and more efficient way of using quantum mechanics to send encrypted information.

Quantum cryptography is more accurately referred to as quantum key distribution. The codes in use today make use of both an encrypted message and a key. The key is a sequence of numbers that is known only to the legitimate sender and receiver of the message—traditionally known as Alice and Bob, respectively—and it is necessary to possess both the key and the encrypted message in order to decode the message. If the key is random and used only once, the code is unbreakable. The problem is, of course, guaranteeing that only the legitimate users know the key. This is where quantum cryptography comes in.

The first quantum key distribution system relied on encoding information in the polarization of single photons². The polarization is related to the internal angular momentum of a photon, and when measured it will assume one of two possible values, which can be identified with 0 and 1. Hence the photon polarization is an example of what is known as a quantum bit, or qubit. Alice can encode the key bit in the photon polarization in many different ways. To extract the information with certainty, Bob,

or an eavesdropper, must know how it was encoded. Bob makes a guess and tells Alice what his guess was. If he was right they have a key bit, if not they throw out the result. An eavesdropper, Eve, who has intercepted the photon, does not know how the key bit was encoded, and she must also guess. If she guesses incorrectly, she will introduce errors into the bits that Alice and Bob share, and by comparing a subset of these bits publicly, they can determine whether an eavesdropper was present or not.

The system constructed by Grosshans *et al.*¹ uses laser pulses containing many photons instead of single photons to carry the information about the key. The coherent pulses can be characterized by two sets of numbers: the average values of the amplitude and phase of the electric field; or the average values of the quadrature components of the electric field. If the electric field is represented by a vector in a plane, the former are equivalent to the polar coordinates of the field vector, and the latter are equivalent to its cartesian coordinates. The quadrature components obey an uncertainty relation that prevents them being accurately measured simultaneously, and, unlike qubits, they can assume a continuous range of values—they are described as continuous variables.

Several groups are investigating quantum continuous variables for effective quantum key distribution. Grosshans *et al.*¹ provide an experimental demonstration that introduces a new procedure—'reverse reconciliation'—by which Alice and Bob handle the quantum key, encoded in continuous variables. In this scheme, Alice encodes information about the key in both quadrature components of a pulse's electric field, and Bob chooses to measure one of them. He then tells Alice which one he measured. Because Eve does not know which quadrature component Bob will measure, she has a problem; if she measures the wrong one she will introduce errors that Alice and Bob will be able to detect by comparing a subset of their key bits. The bonus is that this procedure should remain secure, in principle, even if there are signal losses during transmission: Grosshans *et al.* were able to send key bits securely at

the rate of 75 kilobits per second with a transmission loss of slightly more than 50%; the rate reached 1.7 megabits for lossless transmission.

One of the things that makes quantum cryptography work is that quantum information (that is, information stored in a quantum system) cannot be exactly copied. This is known as the no-cloning theorem³. A consequence is that the most obvious action for an eavesdropper who has managed to intercept a message containing key bits — to make a copy of it and send the original on to the intended party — is not an option. Although it is impossible to construct perfect copies, approximate copies are allowed, but there are limits on how good the copies can be⁴. Grosshans *et al.* have explicitly demonstrated that their quantum cryptographic system is secure against an attack using the best possible coherent-state cloner.

The use of continuous variables, rather

than qubits, in quantum information and computing is an expanding area of research and shows great promise⁵. Until recently, results in this area had been purely theoretical, but with the experimental demonstration of quantum key distribution and teleportation using continuous variables⁶, this field of quantum information has entered the laboratory, and may soon arrive at practical applications. ■

Mark Hillery is in the Department of Physics, Hunter College, City University of New York, 695 Park Avenue, New York, New York 10021, USA. e-mail: mhillery@hunter.cuny.edu

1. Grosshans, F. *et al.* *Nature* **421**, 238–241 (2003).
2. Bennett, C. & Brassard, G. in *Proc. IEEE Conf. Computers, Systems and Signal Processing* 175–179 (IEEE, New York, 1984).
3. Wootters, W. K. & Zurek, W. H. *Nature* **299**, 802–803 (1982).
4. Cerf, N. J., Ipe, A. & Rottenberg, X. *Phys. Rev. Lett.* **85**, 1754–1757 (2000).
5. Braunstein, S. L. & Pati, A. K. *Quantum Information Theory with Continuous Variables* (Kluwer, Dordrecht, in the press).
6. Furusawa, A. *et al.* *Science* **282**, 706–709 (1998).

Evolutionary biology

Splitting in space

Diethard Tautz

Disjunct distributions of closely related species are not necessarily the outcome of passive fragmentation of populations. Instead, they can be the consequence of speciation within a population.

Until recently, the overriding credo for explaining how new species are formed has run as follows: first, a population of organisms splits into several subpopulations; once isolated from other members of their own kind, these subpopulations become adapted to local conditions; so, over millions of years, their descendants evolve into new species. This is 'allopatric speciation', a concept in which spatial separation comes first and genetic divergence follows, and which has dominated biological thinking for many decades. The alternative, 'sympatric speciation', in which new species are created within a single population, has long been seen as a heresy — to the extent that young biologists would risk their careers if they proposed that such a mechanism could occur¹.

Over the past few years, however, modelling work^{2–4} has shown that spatial separation of populations is not a prerequisite for genetic splitting. Doebeli and Dieckmann (page 259 of this issue⁵) now go even further. They propose that spatial separation is a secondary consequence of adaptive genetic divergence under sympatric conditions. In other words, splitting of a population in space can follow genetic splitting within it.

One of the strongest arguments against sympatric speciation, namely that there are no convincing mechanisms for genetic separation in sympatry, has already been addressed in the previous models^{2–4}. These

models solve the problem of preventing gene flow between differently adapted genotypes, a necessity if speciation is to occur, by giving the individuals an active role in choosing their mates. This is called assortative mating or mate choice, and is a well-documented phenomenon in natural populations. One model³ suggests the parallel evolution of ecological adaptations and signals that enable individuals to recognize mating partners with genetic adaptations that are similar to their own. The other two^{2,4} show that the evolution of the signals, and specific mate choice or sexual selection alone, can in themselves lead to genetic splitting.

But although there are field studies that support these models^{6,7}, most biologists still see sympatric splitting only as an interesting exception. This is because there is a second argument in support of allopatry: common experience shows that closely related species are usually spatially separated. If one takes this pattern as a reflection of the process, one inevitably arrives at the conclusion that, although sympatry is possible, allopatry is the norm. But this is exactly the point at which the new work will change the prevailing view.

Doebeli and Dieckmann⁵ base their model on evolutionary branching^{8,9}, which has already shown its usefulness for understanding the sympatric splitting process³. Evolutionary branching describes a process, known as 'disruptive selection', under



100 YEARS AGO

We have received from the director, Captain S. S. Flower, a copy of a handy little guide (with plan) to the Zoological Gardens at Giza, near Cairo. The general introduction is written in English, French and German, and the names of the animals are given in several languages. The issue of this guide may be taken as an indication that the institution under Captain Flower's charge is in a satisfactory and progressive condition.

From *Nature* 15 January 1903.

50 YEARS AGO

British Scientists of the Twentieth Century. By J. G. Crowther. The book gives the impression that it is the work of two authors. There is *A*, the gifted writer with a neat and clear style, and with a wide reading in the literature of science, particularly in the biographies and publications of his characters; and there is *B*, the Marxist philosopher, ever seeking an opportunity to despise British intelligence, British institutions and British theoretical physics. Most of the writing is done by *A*, but *B* always has the last word. Some examples will show how this schizophrenic method of biography works itself out.

J. J. Thomson. *A*: "J. J. was not only a teacher and discoverer, he was a creator in the method and organization of research This world-wide movement of research students to Cambridge was stimulated by the need for science teaching..." *B*: "The cost to mankind as a whole of a leisured life for Newton, Clerk Maxwell and the scores of geniuses who created the subtlety and depth of the Cambridge tradition was great It is impossible to overlook the adolescent, uncultivated, unintellectual aspect of his mind and school".

Rutherford. *A*: "Rutherford departed suddenly in the midst of a healthy, happy, triumphant and glorious life". *B*: "An unconsciously tragic social figure".

Jeans. *A*: "Why was Jeans so successful as a popular writer? Because he was clear, vivid, confident". *B*: "The enormous circulation of 'The Mysterious Universe' owed much to its meretricious style and shallow philosophical thought Somehow or other, though, Jeans extracted £256,000 out of society The character of his writing for the people, revealed the intellectual bankruptcy of the British educated bourgeoisie".

From *Nature* 17 January 1953.

conditions of 'frequency-dependent competition'. In short, this means that in any given population that becomes adapted to a particular ecological niche, there will be increasing competition among those individuals that are best adapted simply because they are the most frequent ones. The consequence is that their genotypes have a lower probability of being transmitted to the next generation than the less frequent genotypes, which use only parts of the resource spectrum of the given niche. This leads to disruptive selection for specialization and consequently to the population's splitting into two new species with differential adaptations.

Doebeli and Dieckmann now add a spatial component to this process by considering an uneven distribution of resources caused, for example, by environmental gradients in temperature, nutrients or altitude. In this situation, local adaptation along the gradient increases the chance that interactions occur between similar individuals, and hence increases the strength of frequency-dependent selection. This leads to two surprising results. First, a sharp geographical separation of populations is generated during the splitting process, although the resource distribution remains continuous. Second, the ecological and genetic conditions under which this occurs are even easier to fulfil than in the previous model without the spatial component. Most interestingly, environmental gradients with intermediate slopes work better than gradients with steep slopes, a result that is in complete contrast to predictions of classical models that are based on allopatric concepts.

Is this all only modellers' fantasy? As yet there is no direct evidence to confirm the predictions of the model, but there are studies pointing in the right direction. For instance, Ogden and Thorpe¹⁰ have looked at the genetic differentiation of lizard populations on the Caribbean island of Martinique. By sampling the populations along carefully controlled transects, they showed that there is a sharp reduction of gene flow along a transect covering habitats at different altitudes, but not along two control transects within homogeneous habitats. Intriguingly, they also find no reduction of gene flow across an old allopatric split, indicating that habitat ecology is more important for gene flow than historical contingencies. Thorpe and Richard¹¹ previously drew a similar conclusion in a comparable study of lizards on the island of Tenerife.

It is satisfying that the inclusion of more realistic conditions into an abstract model of sympatric speciation leads to results that explain natural patterns that have long been used as arguments against sympatric speciation. This shows that the common experience that closely related species are usually spatially separated cannot be taken as direct evidence for the prevalence of allopatric speciation. But then, science has a habit of

showing that common experience is not always a reliable guide to reality. ■

Diethard Tautz is in the Department of Genetics, University of Cologne, Weyertal 121, 50931 Cologne, Germany.
e-mail: tautz@uni-koeln.de

1. Bush, G. L. in *Endless Forms: Species and Speciation* (eds Howard, D. & Berlocher, S.) 425–438 (Oxford Univ. Press, 1998).
2. Kondrashov, A. S. & Kondrashov, F. A. *Nature* **400**, 351–354 (1999).
3. Dieckmann, U. & Doebeli, M. *Nature* **400**, 354–357 (1999).
4. Higashi, M., Takimoto, G. & Yamamura, N. *Nature* **402**, 523–526 (1999).

5. Doebeli, M. & Dieckmann, U. *Nature* **421**, 259–264 (2003).
6. Schlieven, U. *et al. Mol. Ecol.* **10**, 1471–1488 (2001).
7. Lande, R., Seehausen, O. & van Alphen, J. J. *Genetica* **112–113**, 435–443 (2001).
8. Metz, J. A. J., Geritz, S. A. H., Meszina, G., Jacobs, F. J. A. & van Heerwaarden, J. S. in *Stochastic and Spatial Structures of Dynamical Systems* (eds van Strien, S. J. & Verduyn Lunel, S. M.) 183–231 (North Holland, Amsterdam, 1996).
9. Geritz, S. A. H., Kisti, E., Meszina, G. & Metz, J. A. J. *Evol. Ecol.* **12**, 35–57 (1998).
10. Ogden, R. & Thorpe, R. S. *Proc. Natl Acad. Sci. USA* **99**, 13612–13615 (2002).
11. Thorpe, R. S. & Richard, M. *Proc. Natl Acad. Sci. USA* **98**, 3929–3934 (2001).

Developmental biology

Germ-cell attraction

Prabhat S. Kunwar and Ruth Lehmann

Cells must often travel long distances to carry out their assigned tasks in the body. New work reveals how the precursors of eggs and sperm are guided during their epic journey to the gonads.

All sexually reproducing organisms are composed of two types of cells: the 'mortal' somatic cells, which form the body of the organism, and the 'immortal' germ cells, which produce the next generation. During development, precursor germ cells (better known as primordial germ cells, PGCs) are created in one part of the embryo, often far away from their final destination. They must then migrate to the somatic part of the future gonads, where they become mature germ cells — sperm or eggs. Genetic analyses of fruitflies, zebrafish and mice have led to the identification of several genes that are required for PGC migration¹. Moreover, these studies have suggested that the somatic tissues that line the migratory path provide attractive, repulsive and survival cues that direct the germ-cell precursors^{1,2}. Such cues have so far proved elusive. But, writing respectively in *Cell* and on page 279 of this issue, Doitsidou *et al.*³ and Knaut *et al.*⁴ show that a pair of evolutionarily conserved molecules — the protein SDF-1 and its receptor CXCR4 — act to guide PGCs in zebrafish.

Cell migration is important for many biological processes, including embryonic development, immune responses, wound repair and the spread of tumour cells (metastasis). In many cases, proteins known as chemokines have been shown to trigger and guide cell migration. Initially identified as 'defence' proteins because of their ability to regulate immune-cell movement, chemokines have since also been found to be involved in events as diverse as blood and blood-vessel formation and embryo development⁵.

Broadly speaking, these proteins work by interacting with members of a superfamily of receptors found on the surface of motile cells. These are the G-protein-coupled receptors, so called because they interact with intracellular 'on/off' switches known as G proteins.

The interaction between chemokine and receptor activates intracellular signal-transduction cascades, which induce cell movement, survival and gene expression in a manner that depends in part on the precise cell type and signal. For instance, during directed cell migration, the responding cells detect a small concentration gradient of the chemokine, such that the leading edge of the cell orients towards the signal's source. The activated receptor at the front edge stimulates a specific intracellular enzyme, leading to the recruitment of proteins that contain a structural region known as the pleckstrin-homology domain, involved in protein-protein interactions. This is thought to locally amplify the response to the signal⁶.

One chemokine–receptor pair consists of the chemokine SDF-1 and the receptor CXCR4. The receptor was first identified as being important for HIV-1 to enter cells, and the chemokine–receptor duo has also been implicated in the metastasis of breast-cancer cells^{7,8}. Moreover, genetic analysis of these two proteins in mice has shown that they are required for the development of blood cells (for example, in the migration of mature B cells to the bone marrow) and in the central nervous system (in the migration of granule cells within the layers of the cerebellar cortex, for instance)^{9–14}. Doitsidou *et al.*³ and Knaut *et al.*⁴ now show that SDF-1 and CXCR4 are also involved in PGC migration in zebrafish.

The two groups used different techniques to tackle the problem of how PGCs are guided to their destination in zebrafish embryos. Doitsidou *et al.*³ started by inactivating a large group of genes, one at a time, that they thought might be important for PGC migration. To do so, they used a technique known as 'morpholino antisense knockdown'. This led them to *cxc4b*, one of two genes that encode CXCR4 proteins in zebrafish; when

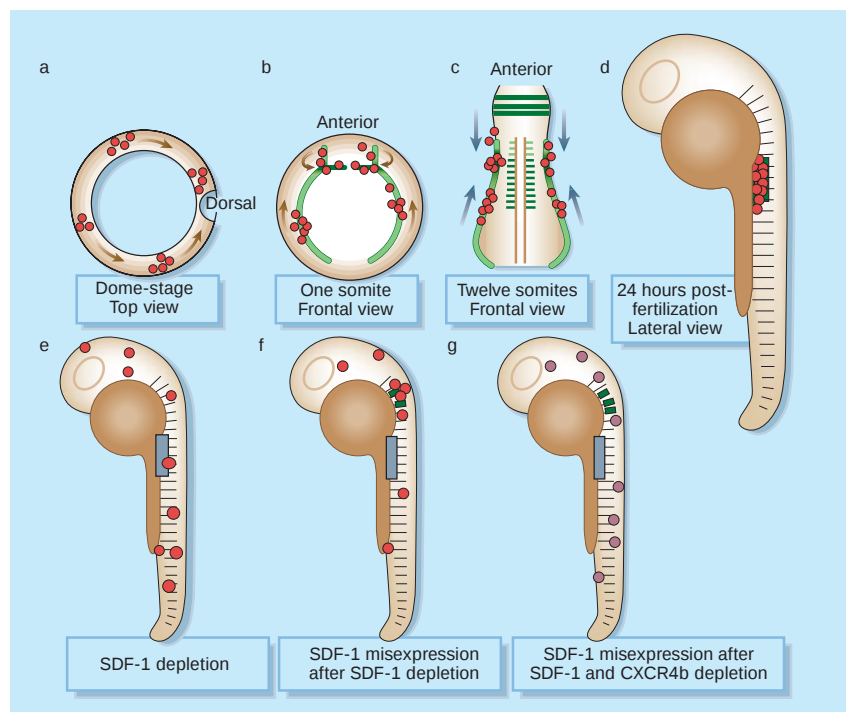


Figure 1 Guiding primordial germ cells (PGCs) in zebrafish. a–d, Normal migration of PGCs (red), incorporating the results of Doitsidou *et al.*³ and Knaut *et al.*⁴. a, Four clusters of PGCs originate at random positions along edges of the embryo and migrate towards the dorsal midline. b, At the beginning of somitogenesis (segment formation), *sdf-1* RNA (green) is expressed strongly near the first segment, and PGCs move towards this location. c, During somitogenesis the four PGC clusters move towards higher levels of *sdf-1*. (Note, however, the high levels of *sdf-1* in the hindbrain — large horizontal stripes — where PGCs do not go.) d, At the end of embryo development PGCs associate with *sdf-1*-expressing cells at the position of the future gonad (green vertical bar). e–g, Studies of SDF-1 and its receptor CXCR4b (refs 3, 4). e, After SDF-1 is depleted (blue–grey vertical bars) or in *ody* mutants (which lack functional CXCR4b; not shown), germ cells scatter throughout the embryo. f, Subsequent misexpression of SDF-1 attracts PGCs to a new position. g, *ody* mutant PGCs (not shown) and CXCR4b-depleted PGCs (purple) do not migrate towards misexpressed SDF-1.

this gene was inactivated, PGCs failed to reach the somatic gonad and instead dispersed throughout the fish. Meanwhile, Knaut *et al.*⁴ made use of a large-scale genetic screen conducted by the Tübingen 2000 Screen Consortium, which randomly generated many thousands of mutant zebrafish. From these mutants Knaut *et al.* identified one, known as *odysseus* (*ody*), in which PGCs likewise failed to reach the gonad. The authors show that the mutation in *ody* animals occurs in the *cxcr4b* gene.

Both groups show that *cxcr4b* is expressed in PGCs, and Knaut *et al.* demonstrate that such expression is necessary for normal PGC migration. Moreover, it appears that when *cxcr4b* is inactivated or mutated, PGCs move, but they cannot read a chemotactic gradient^{3,4}. And although *ody* mutant PGCs can recruit a fluorescently labelled pleckstrin-homology domain to their leading edge, the localization of this domain is unstable and reflects the randomness of the migratory path⁴.

SDF-1 is the major chemokine that binds to CXCR4b, so is it, too, necessary for PGC migration? The new papers suggest that it is: it seems that germ-cell precursors are guided

towards somatic cells that are producing high levels of SDF-1. The migratory path of zebrafish PGCs is complex¹⁵, but each step appears to follow the dynamic expression pattern of SDF-1 (refs 3, 4; Fig. 1a–d). Moreover, alterations in the pattern of SDF-1 expression, resulting from mutation of embryonic tissue-patterning genes, correlate with defects in PGC migration³. And reducing SDF-1 expression produces the same effects as mutating *cxcr4b*. Finally, incorrect expression of SDF-1 at a new location is sufficient to attract PGCs, and this response requires CXCR4b (Fig. 1e–g)^{3,4}.

These studies raise many interesting questions. For instance, are SDF-1 and CXCR4b the only chemokine–receptor pair that guides PGC migration in zebrafish? Possibly not: although PGCs travel to many tissues in which SDF-1 is strongly expressed, they ignore other tissues that express high levels of SDF-1 (Fig. 1c). So, additional signals, modifications of SDF-1, or interactions of SDF-1 with the proteins of the extracellular matrix in which cells are embedded may alter the chemokine's activity as a PGC attractant, and could be needed to

keep these cells on course. Such secondary signals may not be sufficient, however, to attract or repel PGCs, as these cells do not appear to have a favoured destination when the levels of CXCR4b and SDF-1 are decreased.

Also, how does SDF-1 work? Do somatic cells release SDF-1, producing a long-range gradient that attracts PGCs from far away? Or are PGCs instead simply 'caught' when they move close to somatic cells that are expressing SDF-1 on their surface? Which signalling molecules within PGCs mediate the migratory response? And finally, how do germ cells 'know' when they have arrived at their destination? Zebrafish embryos provide an excellent system in which to address these questions.

It will also be interesting to learn whether this chemokine–receptor pair is involved in PGC migration in other organisms. Mice lacking either CXCR4 or SDF-1 have been analysed, but details of PGC migration have not yet been reported. Interestingly, these mice die as embryos with defects in several tissues^{9–14}. By contrast, in zebrafish, the *ody* mutants survive, and the SDF-1–CXCR4b pair seems to be required only during the migration of PGCs and cells known as lateral-line cells^{3,4,16}. Different G-protein-coupled chemokine receptors, such as CXCR4a, may carry out other functions of CXCR4 in zebrafish. In fruitflies, no clear relatives of SDF-1 and CXCR4 have been identified, but G-protein-coupled receptors do seem to be important for PGC migration¹⁷. Indeed, in all organisms studied so far PGCs are seen crawling along, rather like other cells that use signals from G-protein-coupled receptors to trigger movement up chemical gradients⁵. So such signalling may be a major determinant, and the common denominator, of migrating PGCs throughout the animal kingdom. ■

Prabhat S. Kunwar and Ruth Lehmann are at the Howard Hughes Medical Institute, Developmental Genetics Program, Skirball Institute of Biomolecular Medicine, New York University School of Medicine, 540 First Avenue, New York, New York 10016, USA. e-mail: Lehmann@saturn.med.nyu.edu

1. Starz-Gaiano, M. & Lehmann, R. *Mech. Dev.* **105**, 5–18 (2001).
2. Wylie, C. *Cell* **96**, 165–174 (1999).
3. Doitsidou, M. *et al.* *Cell* **111**, 647–659 (2002).
4. Knaut, H. *et al.* *Nature* **421**, 279–282 (2003); advance online publication, 18 December 2002 (doi: 10.1038/nature01338).
5. Horuk, R. *Cytokine Growth Factor Rev.* **12**, 313–335 (2001).
6. Iijima, M., Huang, Y. E. & Devreotes, P. *Dev. Cell* **4**, 469–478 (2002).
7. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. *Science* **272**, 872–877 (1996).
8. Muller, A. *et al.* *Nature* **410**, 50–56 (2001).
9. Hargreaves, D. C. *et al.* *J. Exp. Med.* **194**, 45–56 (2001).
10. Tachibana, K. *et al.* *Nature* **393**, 591–594 (1998).
11. Ma, Q. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 9448–9453 (1998).
12. Zou, Y. R. *et al.* *Nature* **393**, 595–599 (1998).
13. Bagri, A. *et al.* *Development* **129**, 4249–4260 (2002).
14. Zhu, Y. *et al.* *Nature Neurosci.* **5**, 719–720 (2002).
15. Raz, E. *Semin. Cell Dev. Biol.* **13**, 489–495 (2002).
16. David, N. B. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 16297–16302 (2002).
17. Starz-Gaiano, M. *Molecular Guidance Cues and Intracellular Signaling in Drosophila Germ Cell Migration*. Thesis, New York Univ. (2002).

Developmental biology

Decoding stem cells' powers

Genes Dev. **17**, 126–140 (2003)

The therapeutic promise of embryonic stem (ES) cells has dominated headlines — but scientists remain largely ignorant of the genes responsible for their powers. This month, however, an international group of researchers has identified one such gene.

Ariel A. Avilion *et al.* have shown that *Sox2* is essential for the survival of healthy ES cells, which are referred to as 'pluripotent' because they spawn virtually every tissue type in the body. Mouse embryos that had been genetically engineered to lack *Sox2* quickly perished, as did ES cells grown from them. Embryos without *Sox2* could not sustain pluripotent stem cells in the epiblast, the tissue that normally gives rise to the embryo body. Death would occur even earlier, the team suspects, were it not for a dose of *Sox2* protein inherited from the mother's egg.

Sox2 is only the second gene identified as appearing to endow ES cells with pluripotency. The researchers suspect that it works together with the other known gene, called *Oct4*, to switch on a series of essential regulatory molecules.

By teasing apart this genetic hierarchy, it is hoped that researchers will be able to manipulate both embryonic and adult stem cells — and coax them into growing the tissues needed for treatments. **Helen Pearson**

Atmospheric science

Lightning satellite survey

J. Geophys. Res. **108** doi:10.1029/2002JD002347 (2003)

In the most robust survey of lightning activity to date, Hugh Christian and colleagues report that the number of lightning flashes worldwide is less than half the usually accepted estimate. The survey

consists of five years of data from the Optical Transient Detector (OTD), on board the MicroLab-1 satellite, which was specifically designed to spot cloud-to-ground and cloud-to-cloud lightning during night or day. One summary of the authors' results is shown on the map reproduced below.

The new data show that, on average, there are 1.4 billion flashes worldwide each year: in other words, 44 every second. The most trusted previous calculation — now 78 years old, and based on records of thunderstorm-days worldwide — produced an estimate of 100 flashes per second. Given the importance of lightning as a source of nitrogen oxides, for instance, a revision as large as this will interest researchers modelling atmospheric chemistry.

The world's lightning hotspot is near Kamembe, Rwanda, with a mean annual density of 82.7 flashes per square kilometre. More broadly, there is ten times more lightning activity over land than over sea. And 78% of activity occurs between latitudes 30° S and 30° N.

The OTD sensor is now out of service.

Replacing it with an instrument in a geostationary (rather than low-Earth) orbit could provide unique data about the ways in which climatic phenomena such as El Niño and global warming affect lightning activity. **Tom Clarke**

Quantum physics

The entangling effects of motion

Phys. Rev. Lett. **89**, 270402 (2002)

Quantum entanglement, the phenomenon that links the fates of quantum particles by a seemingly instantaneous action at a distance, can be created by motion, according to Robert Gingrich and Christoph Adami.

Entanglement is the key to the burgeoning field of quantum information, which encompasses processes such as quantum teleportation, cryptography and computing. Entangled particles share a

wavefunction, and measurements on one particle automatically define the quantum state of others with which it is entangled.

Gingrich and Adami show that, in a moving frame, quantized properties of two particles (such as spin or momentum) can become entangled when, at rest, they are not. But this weaving together does not appear out of nowhere. It results from the transference of entanglement from one set of quantum variables to another — from spin to momentum, say. If one were measuring only spin, for example, the particles might appear to become magically entwined.

It is a relativistic effect, and represents another step towards the dovetailing of quantum information theory with general relativity. This becomes important when quantum techniques (such as high-precision time-keeping) are deployed in moving frameworks on satellites, for example. And relativistic transfer of entanglement could be useful for preparing and manipulating entangled states in quantum information technologies. **Philip Ball**

Physiology

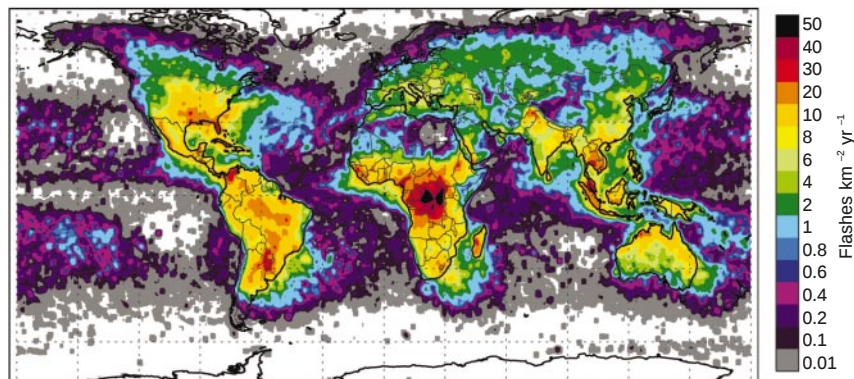
Machinations of the heart's strain gauge

Cell **111**, 943–955 (2002)

To avoid injury, heart muscles must detect and respond to mechanical stress. Ralph Knöll *et al.* have now identified a crucial component of the stress sensor in cardiac muscle, and have shown that it is linked to human disease.

The component in question is a protein called MLP, which lies in the muscle's Z-disk. The Z-disk is the site of attachment of filaments of the actin protein, and has an important structural role in muscle. But it seems that it also has a less passive function. The giant elastic protein titin stretches from one Z-disk to the next, attaching to another protein, telethonin. It now transpires that telethonin also binds MLP, and that in mice genetically modified to lack MLP, the regular structure of the Z-disk breaks down. This suggests that the telethonin–MLP complex detects tension in titin, helping to maintain an ordered muscle structure.

Significant numbers of people suffering from the heart disease dilated cardiomyopathy were found to have a mutation in their MLP gene. The mutation affects a region of the protein that normally binds telethonin, abolishing the interaction. Indeed, the ultrastructure of heart muscle from a patient with dilated cardiomyopathy was very similar to that of heart muscle from MLP-deficient mice. This new information increases our understanding of such heart disorders, but it remains to be seen whether it will translate into new treatments. **Christopher Surridge**



Flash geography — map showing the annual distribution of lightning worldwide. Central Africa experiences the greatest activity. (Image courtesy of H. J. Christian/American Geophysical Union.)

Interaction between Kilauea and Mauna Loa

Last year witnessed an unexpected communication between this pair of volcanoes.

After almost a decade of very slow rates of deformation, Mauna Loa in Hawaii, the largest volcano on Earth, began inflating in May 2002; at the same time, a high-volume effusive episode began at its neighbour Kilauea. We have found a correlation between these events at a very short timescale, detected by continuous deformation monitoring. This remarkable observation suggests that there is a crustal-level interaction between the magma systems of Mauna Loa and Kilauea, reviving a century-old controversy over the relationship between these two volcanoes on the basis of differences in their lava chemistry and in their patterns of eruptive behaviour^{1–5}.

To evaluate the significance of the temporal correlation, we used a Monte Carlo simulation to predict the probability of a rate change at Mauna Loa occurring within one week of an effusive episode at Kilauea during any given 1,000-day interval (about the length of the global positioning system (GPS) time series). Because we used upper-bound estimates of the frequency of such events — three effusive events per year on Kilauea and one deformation-rate change at Mauna Loa every four years — the simulation should yield an upper boundary on the probability of random coincidence, which we find to be less than one in ten. Compared with this figure, the plausibility of a causal interaction between the two volcanoes seems high.

Beyond pure coincidence, how can this correlation be explained? One possibility is a relaxation of stress on Mauna Loa's flank



Figure 2 Summit conference: westward view across the crater of the Hawaiian volcano Kilauea, with Mauna Loa in the background.

that allowed magma to rise to shallow levels. However, because of the small magnitude of the deflation at Kilauea's summit that accompanied the effusive event, it is unlikely that elastic stress transfer could have affected Mauna Loa's magma system significantly.

Another possibility is that there was an increase in magma supply to both volcanoes, contrary both to geochemical models that require mantle-level differences in the parent magmas⁴, and to the suggestion, based on long-term eruptive behaviour, that they compete for magma supply⁵. Also contradictory to this explanation is the lack of rapid inflation immediately before the effusive event. However, in the months leading up to the event, Kilauea's shallow magma system

had been undergoing a slow pressurization (Fig. 1). If this were the first sign of a pulse of magma that was common to both volcanoes, it seems remarkable that it would take a further six months to reach Mauna Loa's shallow system, reaching it by coincidence at the same time as Kilauea produced voluminous flows from a new vent.

Alternatively, the inflation that began in late 2001 at Kilauea may have resulted from reduced lava-transport capacity at the old vent, as it occurred during a period of decreasing surface effusion. Moreover, the extension rate at Kilauea abruptly slowed at the start of the effusive event, turning to contraction in late July, while Mauna Loa continued to extend (Fig. 1).

Although we do not discount the idea of a shared pulse of magma to both volcanoes (Fig. 2), there is a possibility that does not require it. A pulse of magma introduced into Mauna Loa's plumbing system could have increased the pressure in Kilauea's adjacent, and already overpressurized, shallow magma system. The existing instability at Kilauea could have amplified the effect of what is likely to have been a very small perturbation, providing the added stress that was necessary to trigger the effusive episode.

Asta Miklius, Peter Cervelli

US Geological Survey, Hawaiian Volcano Observatory,
Hawaii National Park, Hawaii 96718, USA
e-mail: asta@usgs.gov

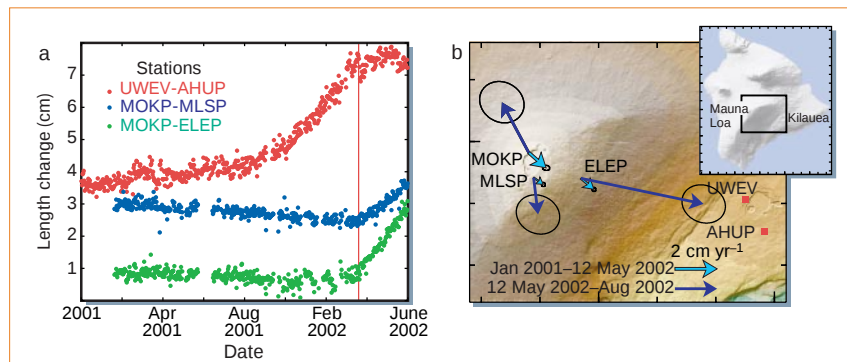


Figure 1 Recent global positioning system (GPS) results from the Mauna Loa and Kilauea volcanoes in Hawaii. **a**, Changes in line length between GPS stations. Time series for stations spanning the summit caldera of Mauna Loa (blue) show the reversal from an almost decade-long trend of slow ($\sim 4 \text{ mm yr}^{-1}$) contraction to extension of more than 5 cm yr^{-1} . The velocity of station ELEP on the south flank also accelerated to more than ten times its previous rate (green). The rate changes coincide with the onset of an effusive event at Kilauea on 12 May 2002 (vertical red line) that emanated from a new vent at the base of Pu'u 'O'o cone on the east rift zone. This event followed a period of inflation at Kilauea (as shown by the extension on the UWEV-AHUP baseline, red) that began in late 2001 and continued, more slowly after 12 May, until late July. **b**, Horizontal-velocity vectors on Mauna Loa (with the motion of the Pacific Plate removed) before and after the 12 May event. The current pattern of deformation at Mauna Loa's summit is consistent with inflation of the magma reservoir, imaged with geodetic data during the volcano's latest eruption, in 1984 (ref. 6). Error ellipses represent 95% confidence regions and are based on repeatability about a constant velocity.

1. Dana, J. D. *Am. J. Sci.* **60**, 235–244 (1850).
2. Jagger, T. A. *Am. J. Sci.* **193**, 255–288 (1917).
3. Rhodes, J. M. *et al. Nature* **337**, 257–260 (1989).
4. Frey, F. A. & Rhodes, J. M. *Phil. Trans. R. Soc. Lond. A* **342**, 121–136 (1993).
5. Klein, F. W. *J. Geophys. Res.* **12**, 1–35 (1982).
6. Lockwood, J. P. *et al. US Geol. Surv. Prof. Pap.* **1350**, 537–570 (1987).

Competing financial interests: declared none.

Tubulin acetylation and cell motility

Although the protein tubulin is known to undergo several post-translational modifications that accumulate in stable but not dynamic microtubules inside cells, the function of these modifications is unknown. Hubbert *et al.*¹ have shown that the enzyme HDAC6 (for histone deacetylase 6) reverses the post-translational acetylation of tubulin, and provide evidence that reducing tubulin acetylation enhances cell motility. They also suggest that decreasing tubulin acetylation reduces microtubule stability. However, we find that microtubule stabilization is not promoted by tubulin acetylation. We conclude that the alteration in cell motility observed by Hubbert *et al.* in cells overexpressing HDAC6 results not from changes in the formation of stable microtubules, but from alterations in the degree of tubulin acetylation.

Most mammalian cells possess two subsets of microtubules: dynamic microtubules with a half-life of 5–10 min, and stable microtubules that have a half-life of hours,

and which contain one or more types of post-translationally modified tubulin². One of these modifications, detyrosination, accumulates in stable microtubules but does not cause microtubule stabilization^{3–7}. For other tubulin modifications, however, the case is less clear.

Hubbert *et al.*¹ did not investigate whether changes in tubulin acetylation alter microtubule stability. To test this, we treated wound-edge, serum-starved NIH 3T3 fibroblasts, which have few stable microtubules^{6,8}, with inhibitors of HDAC6 and used resistance to depolymerization by nocodazole and accumulation of detyrosinated tubulin as assays for increased stable microtubules^{6,8}. Cells treated with trichostatin A (TSA), an inhibitor of HDAC6, showed an increase in microtubule acetylation¹ (Fig. 1a, b, insets), but not in the detyrosination of microtubules compared with untreated cells (Fig. 1a, b). Cells treated with sodium butyrate, a deacetylase inhibitor that does not affect HDAC6 activity¹, did not increase either acetylation or detyrosination of microtubules (results not shown).

Serum-starved cells treated with TSA did not contain nocodazole-resistant microtubules either (Fig. 1c), in contrast to cells treated with a physiological stimulator of stable microtubules, lysophosphatidic acid (LPA)^{6,8} (Fig. 1d). LPA-treated cells had more acetylated microtubules (results not shown).

These results indicate that increased tubulin acetylation does not increase levels of stable microtubules; rather, microtubules must be stabilized by other mechanisms (such as capping⁷) and then these stable microtubules accumulate acetylated tubulin, just as they accumulate detyrosinated tubulin. This is consistent with results showing that tubulin acetylation has no effect on microtubule assembly *in vitro*⁹ and that acetylated tubulin is only detectable in long-lived stable microtubules *in vivo*¹⁰.

Hubbert *et al.* found that HDAC6 overexpression enhances cell motility¹. Our results imply that this increase in cell motility is not caused by changes in levels of stable microtubules, but by changes in the acetylation of tubulin (or of an as-yet-unidentified protein). Migrating wound-edge fibroblasts contain stable, post-translationally modified microtubules that are orientated towards the cell's leading edge^{2,6,8} (Fig. 1d), and these may direct organelles and other important cellular components to the leading edge.

Detyrosinated tubulin seems to have an enhanced affinity for kinesin *in vitro*¹¹, and could be involved in kinesin-dependent

recruitment of intermediate filaments to microtubules¹² and in the recycling of endocytic vesicles¹³. Perhaps acetylation will also turn out to affect the activity of microtubule-associated proteins or motors on microtubules.

Alexander Palazzo*, Brian Ackerman*, Gregg G. Gundersen†*

Departments of *Anatomy and Cell Biology, and †Pathology, Columbia University, New York, New York 10032, USA

e-mail: ggg1@columbia.edu

- Hubbert, C. *et al.* *Nature* **417**, 455–458 (2002).
- Bulinski, J. C. & Gundersen, G. G. *Bioessays* **13**, 285–293 (1991).
- Webster, D. R., Gundersen, G. G., Bulinski, J. C. & Borisy, G. G. *Proc. Natl Acad. Sci. USA* **84**, 9040–9044 (1987).
- Khawaja, S., Gundersen, G. G. & Bulinski, J. C. *J. Cell Biol.* **106**, 141–149 (1988).
- Webster, D. R., Wehland, J., Weber, K. & Borisy, G. G. *J. Cell Biol.* **111**, 113–122 (1990); erratum *ibid.* 1325–1326 (1990).
- Cook, T. A., Nagasaki, T. & Gundersen, G. G. *J. Cell Biol.* **141**, 175–185 (1998).
- Infante, A. S., Stein, M., Zhai, Y., Borisy, G. & Gundersen, G. G. *J. Cell Sci.* **113**, 3907–3919 (2000).
- Palazzo, A. F., Cook, T. A., Alberts, A. S. & Gundersen, G. G. *Nature Cell Biol.* **3**, 723–729 (2001).
- Maruta, H., Greer, K. & Rosenbaum, J. L. *J. Cell Biol.* **103**, 571–579 (1986).
- Webster, D. R. & Borisy, G. G. *J. Cell Sci.* **92**, 57–65 (1989).
- Liao, G. & Gundersen, G. G. *J. Biol. Chem.* **273**, 9797–9803 (1998).
- Kreitzer, G., Liao, G. & Gundersen, G. G. *Mol. Biol. Cell* **10**, 1105–1118 (1999).
- Lin, S. X., Gundersen, G. G. & Maxfield, F. R. *Mol. Biol. Cell* **13**, 96–109 (2002).

correction

Visual structure of a Japanese Zen garden

G. J. Van Tonder, M. J. Lyons, Y. Ejima

Nature **419**, 359–360 (2002)

In the legend for Fig. 2, the date AD 1681 is incorrect: in fact, the plan of the garden and temple indicates their likely layout before the building was destroyed by fire in AD 1797 and is based on ref. 4 of our communication. This error does not affect our conclusions.

addendum

Magnetic shape-memory effects in a crystal

A. N. Lavrov, S. Komiya, Y. Ando

Nature **418**, 385–386 (2002)

It has been drawn to our attention that the magnetic shape-memory effects we reported in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ (LSCO) crystals bear similarities to the conventional magnetostriction associated with antiferromagnetic domain structures. Indeed, in the Néel state, static antiferromagnetic domains may generate in LSCO crystals a pattern of structural distortions that can be modified by magnetic fields. However, we find that the magnetic shape memory in LSCO is a distinct phenomenon whereby magnetic fields affect genuine orthorhombic domains in both antiferromagnetic and paramagnetic states of LSCO, regardless of the existence of a magnetic order. This was not made sufficiently clear in our communication.

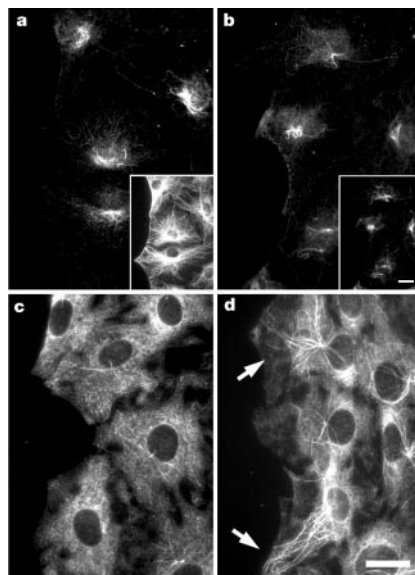


Figure 1 Immunofluorescent images of serum-starved, wounded NIH 3T3 fibroblasts, showing that increased acetylation of tubulin does not stabilize microtubules in these cells. Cells were incubated with trichostatin A (TSA; 5 μM ; 4 h; **a, c**) or without TSA (**b, d**); cells in **d** were treated with 10 μM lysophosphatidic acid (LPA). Cells were fixed and immunostained for detyrosinated tubulin^{6,8} (**a, b**), acetylated tubulin¹⁰ (insets) or bulk tubulin (**c, d**). **a, b**, TSA increases microtubule acetylation (insets) but does not increase microtubule detyrosination compared with untreated controls (**a, b**). **c, d**, TSA does not increase the number of microtubules that are resistant to nocodazole (**c**; 10 μM ; 30 min), whereas cells treated with LPA have nocodazole-resistant microtubules (**d**). Arrows show stable, modified microtubules orientated towards the leading edge. Scale bars, 15 μm .

brief communications is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on *Nature's* website (www.nature.com/nature) or on request from nature@nature.com.

Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi

Ravi S. Kamath^{*†}, Andrew G. Fraser^{*†§}, Yan Dong^{*}, Gino Poulin^{*}, Richard Durbin[‡], Monica Gotta^{*§}, Alexander Kanapin^{||}, Nathalie Le Bot^{*}, Sergio Moreno^{*¶}, Marc Sohmann^{‡§}, David P. Welchman^{*}, Peder Zipperlen^{*} & Julie Ahringer^{*}

^{*} Wellcome Trust/Cancer Research UK Institute and Department of Genetics, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK

[‡] Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

^{||} EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

[¶] Centro de Investigación del Cáncer, CSIC / Univ. Salamanca, Campus Miguel de Unamuno, 37007 Salamanca, Spain

[†] These authors contributed equally to this work

A principal challenge currently facing biologists is how to connect the complete DNA sequence of an organism to its development and behaviour. Large-scale targeted-deletions have been successful in defining gene functions in the single-celled yeast *Saccharomyces cerevisiae*, but comparable analyses have yet to be performed in an animal. Here we describe the use of RNA interference to inhibit the function of ~86% of the 19,427 predicted genes of *C. elegans*. We identified mutant phenotypes for 1,722 genes, about two-thirds of which were not previously associated with a phenotype. We find that genes of similar functions are clustered in distinct, multi-megabase regions of individual chromosomes; genes in these regions tend to share transcriptional profiles. Our resulting data set and reusable RNAi library of 16,757 bacterial clones will facilitate systematic analyses of the connections among gene sequence, chromosomal location and gene function in *C. elegans*.

The ability to inactivate a target gene transiently by RNAi¹ has greatly accelerated the analysis of loss-of-function phenotypes in *C. elegans* and other organisms. Although several large-scale RNAi-based screens have been used to study gene function in *C. elegans*^{2–4}, in total only about a third of the predicted genes have been analysed so far. Genome-wide RNAi analyses would not only provide a key resource for studying gene function in *C. elegans* but should also address important issues in functional genomics, such as the global organization of gene functions in a metazoan genome. In addition, because more than half of the genes in *C. elegans* have a human homologue, this kind of functional analysis in the worm should provide insights into human gene function.

Analysis of gene functions by RNAi

Loss-of-function RNAi phenotypes can be generated efficiently by feeding worms with bacteria expressing double-stranded RNA (dsRNA) that is homologous to a target gene^{5–7}; we previously used this method to screen roughly 87% of predicted genes on chromosome I of *C. elegans* (ref. 2). To screen most of the predicted genes in *C. elegans* by RNAi, we constructed a library of bacterial strains, each capable of expressing dsRNA designed to correspond to a single gene. The library consists of 16,757 bacterial strains, which in total correspond to about 86% of the 19,427 current predicted genes in *C. elegans* with similar coverage across each chromosome (see Supplementary Tables 1 and 2). Using this library, we screened wild-type *C. elegans* hermaphrodites to identify genes for which RNAi reproducibly results in sterility, embryonic or larval lethality, slow post-embryonic growth, or a post-embryonic defect (Methods). Such phenotypes were obtained with 1,722 bacterial strains (10.3% of those analysed; Fig. 1a).

Many strains gave rise to several reproducible RNAi phenotypes, indicating that the targeted gene has many developmental roles. For example, RNAi against Y77E11A.13a (which encodes a homologue

of the yeast Sec13p protein implicated in protein trafficking from the endoplasmic reticulum to the Golgi⁸) results in sterility, embryonic lethality or uncoordinated movement. To simplify subsequent genomic analyses, we defined three mutually exclusive phenotypic classes: the nonviable (Nonv) class, consisting of embryonic or larval lethality or sterility (with or without associated post-embryonic defects); the growth defects (Gro) class, consisting of slow or arrested post-embryonic growth; and the viable post-embryonic phenotype (Vpep) class, consisting of defects in post-embryonic development (for example, in movement or body shape) without any associated lethality or slowed growth. The RNAi phenotypes obtained on each chromosome are summarized in Fig. 1a, and a full list of phenotypes by gene is given in Supplementary Tables 2–4; these data are available publicly on Wormbase (<http://www.wormbase.org>).

To determine the effectiveness of the screen, we assessed our ability to identify correctly the known loss-of-function phenotypes for previously studied loci. Overall, we obtained RNAi phenotypes for 63.5% of 323 detectable loci; almost all of those detected (92%) produced an RNAi phenotype similar to the known mutant phenotype (see Supplementary Tables 5 and 6). More loci with a Nonv phenotype were detected (77.9%) than loci with a Vpep phenotype (42.2%). This difference is likely to arise because certain classes of gene with Vpep phenotypes (for example, neuronally expressed genes) are relatively resistant to RNAi^{7,9} and because Vpep phenotypes are more difficult to detect in this screen (Methods). Notably, the estimated rate of false-positive RNAi phenotypes is very low (<1%; see Supplementary Fig. 1). In addition, our results correlate well with, and are as sensitive as, previous RNAi screens (refs 3, 4, and Supplementary Fig. 1), indicating that RNAi data are highly reproducible irrespective of the method used.

The most common RNAi phenotype is embryonic lethality, which was observed for 929 strains (5.5%). On the basis of our efficiency of detecting known embryonic lethal loci, this probably includes over 70% of embryonic lethal genes and thus will be an excellent starting point for more detailed analyses of the molecular

[§] Present addresses: Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK (A.G.F.); Institute of Biochemistry ETH Hoenggerberg, 8093-Zurich, Switzerland (M.G., M.S.).

mechanisms of embryogenesis in *C. elegans*. Of the post-embryonic phenotypes detected, the largest class was uncoordinated movement (Unc), which is typically indicative of a defect in the neuromuscular system. We also defined an RNAi phenotype for 33 close homologues (BlastP *E* values less than 10⁻⁶) of human disease genes (Table 1). Notably, many of these genes had Vpep phenotypes (50% versus 16% among all genes with a phenotype), consistent with their embryonic viability in humans, and thus may be useful for establishing *C. elegans* models of some human diseases.

A small percentage of the bacterial strains were predicted to target more than one predicted gene. Before carrying out global analyses, we removed these ambiguous data to generate a set of 1,528 clones for which RNAi phenotypes could be attributed to a single predicted gene (Methods).

Conservation and gene function

We and others have previously found relationships between the RNAi phenotype of a gene and its degree of conservation and putative molecular function, using relatively small datasets^{2-4,10}. Using the larger dataset obtained here, we have confirmed and extended those conclusions. We find that *C. elegans* genes with an orthologue in another eukaryote are much more likely to have a detectable RNAi phenotype than all other genes (21% versus 6%).

In addition, highly conserved genes that are present as a single copy in the *C. elegans* genome are more than twice as likely to have an RNAi phenotype as those that are present in more than one copy (31% versus 12%); this suggests that many recently duplicated paralogues are at least partially functionally redundant or have specialized functions that are not detectable in this screen.

The highest cross-species conservation is seen among genes with a Nonv RNAi phenotype, of which 52% have an orthologue in another eukaryote; this shows that similar essential basal cellular machinery is common to all eukaryotes. Indeed, 51% of *C. elegans* orthologues of yeast essential genes¹¹ have a Nonv RNAi phenotype. Consistent with these findings, genes involved in the basic metabolism and maintenance of the cell are significantly enriched for having a Nonv RNAi phenotype (Fig. 2a); by contrast, genes involved in more complex processes that are expanded in metazoa, such as signal transduction and transcriptional regulation, are enriched for Vpep phenotypes (Fig. 2b).

Domain evolution and gene function

To study further the relationship between the sequence and function of a gene, we examined the domain composition of genes in each phenotypic class. Of the 200 most abundant InterPro domains¹² in the *C. elegans* genome, 28 show significant (*P* < 0.05) associations

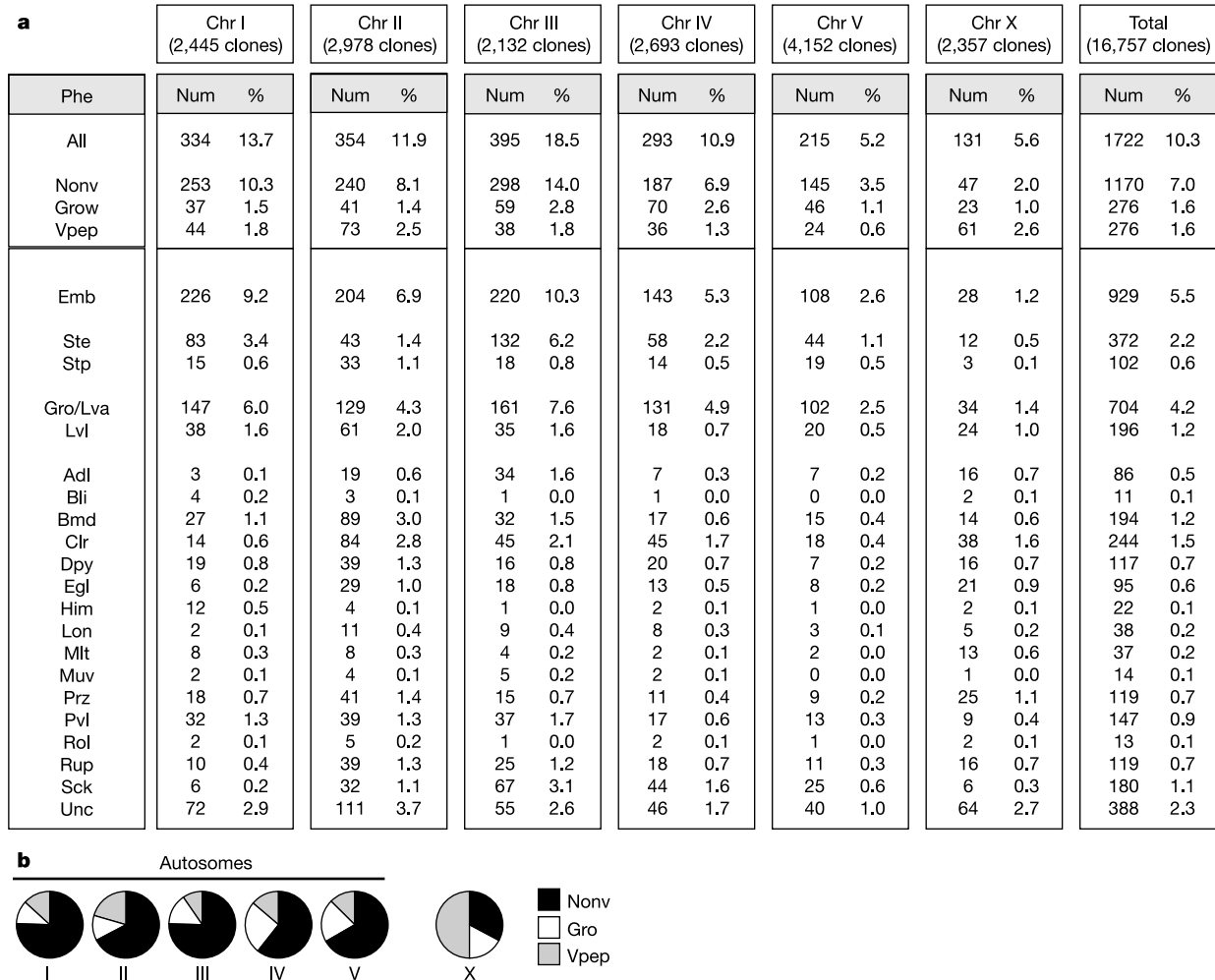


Figure 1 Summary of RNAi phenotypes. **a**, Number of bacterial strains associated with each RNAi phenotype. The Nonv (nonviable, including all phenotypic classes that result in lethality or sterility), Gro (growth defects, including slow post-embryonic growth or larval arrest) and Vpep (viable post-embryonic phenotype, including all other phenotypic

classes) categories are mutually exclusive; however, many genes are associated with several specific RNAi phenotypes. Phenotypic classes are described in Methods. The percentages are out of the total number of clones screened per chromosome. **b**, Relative proportion of Nonv, Gro and Vpep phenotype on each chromosome.

with particular classes of RNAi phenotype (Table 2). Notably, of the seven InterPro domains that are significantly associated with Vpep RNAi phenotypes, most (six) are represented in the fly¹³ and human^{14,15} genomes but not in the genome of budding yeast¹⁶ or *Arabidopsis*¹⁷. Genes with a Vpep phenotype by definition have no associated lethality but instead have a role in the multicellular animal (such as in movement or body shape); therefore, these data suggest that many of the 'animal-specific' functions encoded by genes with Vpep phenotypes may have arisen through the evolution of new domains.

To explore this idea further, we examined whether genes with animal-specific domains are, in general, more likely to have an 'animal-specific' function (that is, to have a Vpep RNAi phenotype). *C. elegans* genes encoding at least one identifiable domain were split into three groups: 'ancient', in which all encoded protein domains are found in yeast, *Arabidopsis*, *Drosophila* and humans; 'animal', in which at least one domain is found in *Drosophila* or humans but not in yeast or *Arabidopsis*; and 'worm', in which any domain is found only in *C. elegans* (37% are ancient, 8% are animal, 10% are worm and 46% have no identifiable domain).

Whereas genes with a Nonv RNAi phenotype are highly enriched for being in the ancient class (Fig. 3; 90% of those with an identifiable domain are 'ancient'), genes with a Vpep RNAi phenotype are enriched for being in the animal class (16% of Vpep genes but only 6% of Nonv genes are in the animal class). This supports the idea that the evolution of new domains has been important for the evolution of animal-specific gene functions. In addition, we found that almost none of the genes in the 'worm' class has an essential role in *C. elegans*, although many have a Vpep phenotype. This suggests that these genes have nematode-specific developmental functions and supports the view that the basal machinery of eukaryotes is shared and not phylum-specific.

The X chromosome

The *C. elegans* genome is organized into five autosomes and a sex

chromosome (X)¹⁸. Sex in *C. elegans* is determined by the number of copies of the X chromosome: hermaphrodites have two copies of the X chromosome, each of which is partially transcriptionally silenced to ensure dosage compensation to and males have a single copy (reviewed in ref. 19). We explored whether there are functional differences between genes on the autosomes and the X chromosome. We found that whereas the autosomes each have a similar distribution of RNAi phenotypes, the distribution on the X chromosome is markedly different (Fig. 1b). This difference is due almost completely to a reduction in the percentage of genes with a Nonv phenotype (Fig. 1a), an effect previously reported by other groups using smaller datasets^{3,10}. Thus, there has been strong selection against the encoding of essential functions on the X chromosome.

Previous studies have shown that X-linked genes are transcriptionally silenced in the germ line during mitosis and early meiosis^{20,21}. Genes required for the basic cellular processes that are essential for the viability of all cells (including those in the germ line) might thus be expected to be absent from the X chromosome; many such genes have Nonv RNAi phenotypes. We indeed found that genes in the functional classes enriched for Nonv phenotypes (such as protein synthesis) are highly underrepresented on the X chromosome (Fig. 2c and Supplementary Fig. 2). The reduction in the number of essential functions encoded on the X chromosome therefore seems to be related to the transcriptional repression of X-linked genes in the germ line. Differential expression of X-linked genes does not explain the entire difference, however, because X-linked and autosomal genes with similar germline expression profiles have very different roles. For example, although genes with oocyte-enriched expression are found in similar numbers on the X chromosome and the autosomes²⁰, none of the X-linked oocyte-enriched genes have a Nonv RNAi phenotype, whereas 19% of the autosomal oocyte-enriched genes are essential.

A second, more intriguing property of the X chromosome is that it is enriched for genes with Vpep phenotypes ($P < 0.01$; chromo-

Table 1 Thirty-three human disease gene homologues with an RNAi phenotype

Predicted gene	<i>C. elegans</i> locus	Human disease	Human gene	BlastP <i>E</i> value	RNAi phenotype
B0035.5		G6PD deficiency	G6PD	1×10^{-176}	Emb, Clr, Gro
B0350.2A	<i>unc-44</i>	Hereditary spherocytosis	ANK1	0.00	Slu
C01G6.8	<i>cam-1/kin-8</i>	Insulin-resistant diabetes mellitus	INSR	6×10^{-55}	Unc, Pvl, clear patch
C01G8.5A		Neurofibromatosis	NF2	1×10^{-123}	Unc, Lvl, Gro
C06A1.1		Zellweger syndrome	PEX1	3×10^{-67}	Emb, Bmd, Sck, Gro
C07H6.7	<i>lin-39</i>	MODY, type IV	IPF1	5×10^{-14}	Egl, Vul, Muv
C17E4.5		Oculopharyngeal muscular dystrophy	PABPN1	3×10^{-41}	Emb, Unc, Lva
C29A12.3	<i>lig-1</i>	DNA ligase I deficiency	DNA ligase1	1×10^{-167}	Emb
C48A7.1	<i>egl-19</i>	Long QT syndrome 3	SCN5A	2×10^{-64}	Egl, Clr
C50H2.1		Leydig cell hypoplasia	LHCGR	9×10^{-76}	Gro
D2045.1		Spinocerebellar ataxia 2	SCA2	7×10^{-09}	Emb
F01G10.1		Wernicke-Korsakoff syndrome	TKT	0.00	Emb, Clr, Gro
F07A5.7	<i>unc-15</i>	Tuberous sclerosis	TSC1	1×10^{-07}	Unc, Prz, Egl
F11C1.6	<i>nhr-25</i>	Pseudohyperaldosteronism	NR3C2	7×10^{-24}	Unc, Prz, Clr, Egl
F11H8.4	<i>cyk-1</i>	Nonsyndromic sensorineural deafness	DFNA1	9×10^{-49}	Emb, Adl, Rup, Clr
F20B6.2	<i>vha-12</i>	Renal tubular acidosis	ATP6B1	0.00	Emb, Ste, Adl, Lvl, Prz
F54D8.1		Ehlers-Danlos syndrome, type IV	COL3A1	1×10^{-06}	Dpy
F53G12.3		Chronic Granulomatous Disease	X-CGD	3×10^{-34}	Bli, Mlt, Lvl
F58A3.2A	<i>egl-15</i>	Multiple venous malformations	VMCM	1×10^{-62}	Egl
K04G2.8A	<i>apr-1</i>	Adenomatous polyposis of the colon	APC	9×10^{-34}	Unc, Bmd, Lvl
K07A1.12	<i>rba-2</i>	Cockayne syndrome	CKN1	6×10^{-13}	Emb, Pvl, Lvl
K08A8.2		Gonadal dysgenesis	SRY	3×10^{-31}	Unc, Egl
K08C7.3	<i>epi-1</i>	Usher syndrome 2a	USH2A	1×10^{-112}	Ste, Unc, Muv, Dpy, Pvl, Rup
K11D9.2A		Darier-White disease	SERCA	0.00	Ste, Sck
M02A10.2		Hyperinsulinism	KCNJ11	4×10^{-78}	Unc
R107.8	<i>lin-12</i>	Alagille syndrome	JAG1	2×10^{-90}	Egl
R12B2.1	<i>sma-4</i>	Pancreatic carcinoma	MADH4	2×10^{-39}	Sma, Dpy
T03F6.5	<i>lis-1</i>	Miller-Dieker lissencephaly syndrome	PAF	1×10^{-148}	Emb
W05E10.3	<i>ceh-32</i>	Holoprosencephaly	SIX3	1×10^{-69}	Unc
W10G6.3	<i>ifa-2</i>	Keratoderma	KRT9	7×10^{-26}	Unc, Lvl, Mlt
Y47D3A.6A	<i>tra-1</i>	Grieg cephalopolysyndactyly syndrome	GLI	6×10^{-58}	Rup, clear patch
Y76A2A.2		Menkes disease	ATP7A	0.00	Prz, Adl, Unc
ZC506.4	<i>mgl-1</i>	Hypercalcemia	CASR	2×10^{-77}	Gro

C. elegans genes with a human disease gene homologue are defined as those with a BlastP *E* value less than 1.0×10^{-6} , taken from refs 38, 39. Shown are those with an RNAi phenotype. The phenotypes are defined in Methods. MODY, maturity onset diabetes of the young. G6PD, glucose-6-phosphate dehydrogenase.

some II is also enriched for Vpep genes). In addition, significantly ($P < 0.01$) more X-linked genes than autosomal genes encode components of signalling pathways and transcription factors; these genes are enriched for Vpep phenotypes. This concentration of Vpep genes on the X chromosome may have evolutionary benefits. Whereas a hermaphrodite worm that is heterozygous for a mutant allele of an X-linked gene is likely to be phenotypically wild type, a (hemizygous) male inheriting the mutant allele will be

mutant. Hermaphrodites could thus act as wild-type repositories for mutant alleles of genes affecting the patterning, structure or behaviour of worms; these alleles could then be selected for or against in a dominant manner in the hemizygous male animal. Because the number of males spontaneously arising from hermaphrodites through meiotic non-disjunction events increases markedly under stressful conditions (such as increased temperature), this haploselection for relatively subtle phenotypic changes might be a powerful mechanism by which to adapt to a changing environment.

Large-scale functional gene clustering

Our RNAi experiments targeted most of the genes in *C. elegans*, with similar proportions of genes covered along each chromosome. Using these data, we examined whether genes of similar function cluster in specific regions of chromosomes. Unlike most animals, *C. elegans* has holocentric chromosomes that lack a localized centromeric region. The five autosomes have a central ‘cluster’, where rates of recombination are low and where most studied genetic loci are found, which is flanked by chromosome ‘arms’, where recombination rates are more than tenfold higher²². These clusters have characteristic features on all autosomes: lower repeat content, greater conservation and greater representation by expressed sequence tags (ESTs)¹⁸. By contrast, the X chromosome does not have a defined cluster region.

In agreement with data derived from classical genetics, we found that genes with RNAi phenotypes are enriched twofold in the cluster regions relative to the arms (7.6% of genes on arms have an RNAi phenotype versus 14.9% in the cluster regions; Fig. 4a). We next examined the distribution of the Nonv, Gro and Vpep genes in the genome (Methods). Notably, genes with a Nonv RNAi phenotype are strongly enriched in large regions of the clusters of chromosomes I, II and III ($P < 0.01$; Fig. 4b): 36% of the Nonv genes lie in

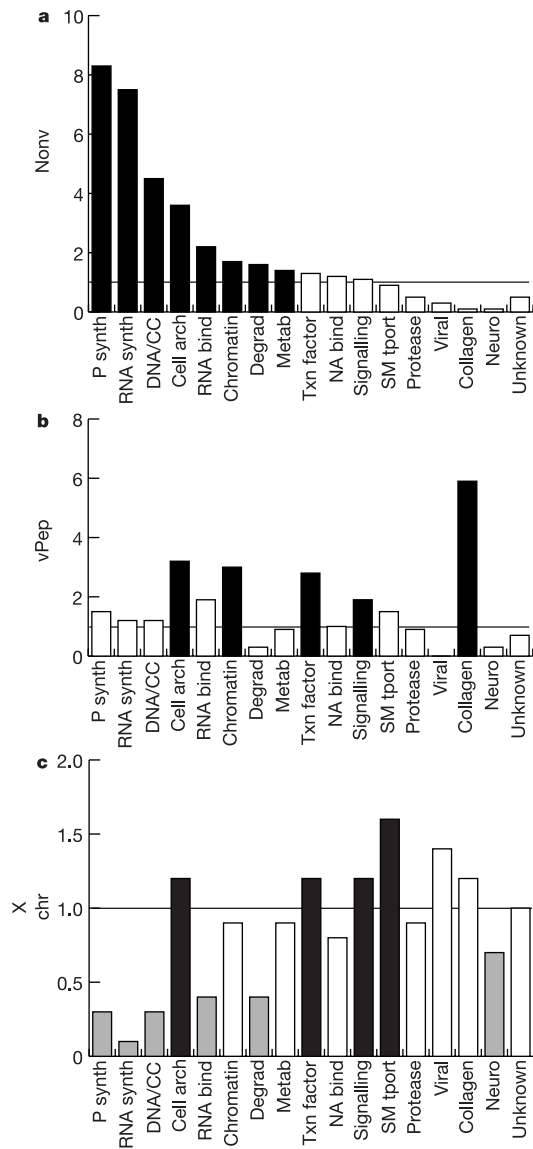


Figure 2 Relative enrichment of Nonv, Vpep and X chromosome genes for different functional classes. The functional classes are protein synthesis (P synth), RNA synthesis (RNA synth), DNA synthesis and repair/cell cycle (DNA/CC), cellular architecture (Cell arch), RNA binding (RNA bind), chromatin regulation (Chromatin), protein degradation (Degrad), energy and intermediary metabolism (Metab), transcription factors (Txn factor), nucleic-acid binding (NA bind), signal transduction (Signalling), small-molecule transport (SM tport), specific proteases (Protease), retroviral- and transposon-derived sequences (Viral), collagens (Collagen), genes with neuronal functions (Neuro), and Unknown. Shown are the levels of enrichment among genes in each functional class for Nonv phenotypes (a), Vpep phenotypes (b) or genes on the X chromosome (c); bars in black denote a statistically significant overenrichment ($P < 0.01$). The grey bars in c represent an underenrichment ($P < 0.01$). For reference, a line is drawn at a relative representation of 1.0.

Table 2 InterPro domains associated with RNAi phenotypes

Nonv only
Elongation factor, GTP-binding
Cyclin
Ubiquitin domain
TPR repeat
Zinc-finger, CCHC type
Myb DNA-binding domain
Laminin-type EGF-like domain
DEAD/DEAH box helicase
Ubiquitin-associated domain
Zinc-finger, C ₂ H ₂ type
Mitochondrial substrate carrier
Protein kinase C, phorbol ester/DAG binding
Gro only
Glycosyl transferase, family 2
Zinc-finger, RING
Phosphotyrosine interaction domain
Proline-rich extensin
Nonv and Gro
G-protein β-subunit WD40 repeat
AAA ATPase
KH domain
Zinc-finger, C-X ₆ -C-X ₅ -C-X ₃ -H type
RNA-binding region RNP-1 (RNA recognition)
Vpep
Immunoglobulin/major histocompatibility complex
Collagen triple helix repeat
Immunoglobulin-like
EGF-like calcium-binding
Aspartic acid and asparagine hydroxylation site
Fibronectin, type III
Worm-specific repeat type 1

We examined the phenotypes of genes containing any of the 200 most abundant InterPro¹² domains in the *C. elegans* genome; genes containing the listed domains were significantly enriched ($P < 0.05$) for the indicated phenotypes, in order of decreasing significance. DAG, diacylglycerol; EGF, epidermal growth factor.

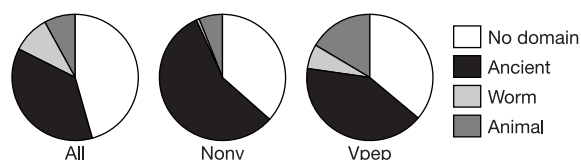


Figure 3 Conservation of domains in genes with different RNAi phenotypes. All predicted genes were placed into one of four mutually exclusive classes on the basis of their InterPro domain content. The ‘ancient’ class comprises genes for which all predicted domains are also encoded in the *S. cerevisiae*, *A. thaliana*, *D. melanogaster* and *H. sapiens* genomes; the ‘animal’ class comprises genes that contain any domain present in the *D. melanogaster* or *H. sapiens* genomes, but not in *S. cerevisiae* or *A. thaliana*, and the ‘worm’ class comprises genes containing any domain present in the *C. elegans* genome, but not in the other four. The proportions of All, Nonv and Vpep genes that fall into each class are shown.

these enriched regions, which represent about 13% of the genome. By contrast, Nonv genes are underenriched on the autosomal arms and the whole of the X chromosome. Functional redundancy among paralogous genes might explain some of the underenrichment, because these regions frequently overlap those areas of the autosomes with increased gene duplication (Fig. 4b).

Genes with Vpep and Gro phenotypes are enriched in different regions of the genome from those showing enrichment for Nonv genes. Notably, genes with a Vpep phenotype are enriched significantly in the centre of the X chromosome, despite the absence of a recombinationally defined cluster²². This suggests that the X chromosome, like the autosomes, has a central accumulation of genes with nonredundant functions; on the X chromosome, however, these genes are not required for viability, but rather for worm behaviour or morphology. These findings suggest that in *C. elegans* there is selective pressure for genes with similar organismal

functions to be colocalized in large domains of the genome.

How such domains are maintained and what they represent mechanistically are unclear. A possible hypothesis is that, perhaps as a consequence of long-range chromatin regulation, genes in these domains are transcriptionally co-regulated. To investigate this possibility, we examined sets or ‘mounts’²³ of *C. elegans* genes identified by microarray analysis to share expression profiles; we found that genes in each mount are enriched in distinct regions of the chromosomes (Supplementary Fig. 3). Such large-scale clustering has also been observed in both humans²⁴ and *Drosophila*²⁵.

Notably, genes in mounts 7 and 11 are significantly enriched in the same regions of the genome as are the Nonv genes (Fig. 4b and Supplementary Fig. 3); in addition, these mounts are enriched for genes with Nonv RNAi phenotypes. This suggests that in regions of the genome that have concentrations of genes of similar functions, there is large-scale broad transcriptional co-regulation. The scale of these regions (over 1 megabase) indicates that this mode of regulation is clearly distinct from that previously reported in yeast²⁶ and in *C. elegans*²⁷, in which small clusters of nearly adjacent genes are likely to be co-regulated, perhaps as a consequence of open loops of chromatin^{26,28}. When an assembled genome sequence is available for the nematode *Caenorhabditis briggsae*, which is closely related to *C. elegans*, it will be intriguing to see whether these functional domains are maintained as syntenic regions.

In summary, we note that there are differences in gene function between the X chromosome and the autosomes, as well as functional clustering in different regions of the genome. Each chromosome has unique features—for example, chromosome V has few essential genes relative to the other autosomes and has a high degree of gene duplications, whereas chromosome III is enriched for Nonv genes, and chromosome II is enriched for Vpep genes. These data suggest that different chromosomes and regions of the genome may be specialized for particular functions.

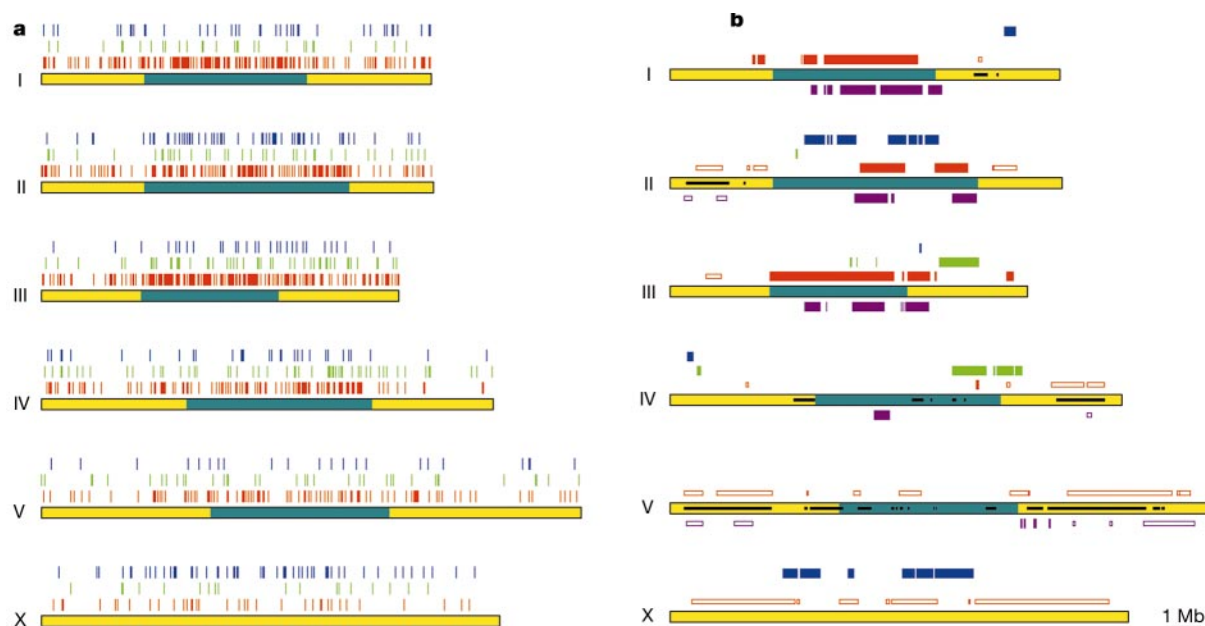


Figure 4 Distribution of RNAi phenotypes across the *C. elegans* chromosomes. **a**, Genomic locations of genes with RNAi phenotypes. Horizontal yellow (arm regions) and blue-green (cluster regions) bars represent *C. elegans* chromosomes; black bars indicate regions enriched for duplicated genes (that is, those with a *C. elegans* homologue). Each RNAi phenotype is represented by a single red (Nonv), green (Gro) or blue (Vpep) line above the chromosomes. **b**, Chromosomal enrichment of genes with different RNAi

phenotypes. Overenrichment is indicated by filled boxes, underenrichment by open boxes. No windows could be significantly underenriched for Gro or Vpep phenotypes owing to the smaller sample sizes. The purple bars below the chromosomes represent regions that are significantly ($P < 0.01$) over- or underenriched for genes in mount 11 (ref. 23). In the enriched regions, 36% of Nonv genes lie in 13% of the genome, 11.6% of Gro genes lie in 3.9% of the genome, and 23.9% of Vpep genes lie in 7.8% of the genome.

Conclusion

We have used RNAi to examine the loss-of-function phenotypes of about 86% of predicted genes in *C. elegans*. To our knowledge, this is the first systematic functional analysis of a metazoan genome. Of the 1,528 genes for which we could assign an RNAi phenotype, over two-thirds had not been previously associated with a biological function *in vivo*. In addition, we have created an RNAi feeding library of bacterial clones that can be replicated and reused for an unlimited number of future genome-wide RNAi screens in *C. elegans*.

Much as the genome sequence has provided an invaluable platform for investigating *C. elegans* biology, these data and the availability of this library will form a useful tool for functional genomic studies in *C. elegans*. In the future, an analogous genome-wide RNAi library approach could be extended to mammalian cells by capitalizing on techniques using DNA constructs to encode hairpin RNAs^{29–34}. We anticipate that in the coming years the quantity of functional data derived from RNAi-based screens in *C. elegans* and in other organisms will greatly expand our understanding of how genes function to bring about the phenotype of an organism. □

Methods

Generation of bacterial feeding library

Polymerase chain reaction (PCR) products were generated using the Research Genetics *C. elegans* GenePairs primer set of 19,213 primer pairs. The set of predicted genes used includes only those genes thought to encode proteins. Primer sequences are listed on the Kim Lab website at Stanford University (<http://cmgm.stanford.edu/~kimlab/primers.12-22-99.html>). Current alignments of predicted GenePair PCR products on the *C. elegans* genome are available at WormBase (<http://www.wormbase.org>). We generated PCR products and constructed bacterial strains as described². Inserts were checked for the correct size and confirmed by PCR using the original GenePair oligomers. The whole-genome library consists of 16,757 clones, which represent 87.2% of the GenePairs set and are predicted to correspond to 86.3% of *C. elegans* predicted genes¹⁸, exclusive of cross-RNAi interactions (see below). To assess the quality of the cloning procedure, we sequenced 100 random clones and found all of them to be correct. For the 13% of GenePairs for which no bacterial strain was made, either the GenePair failed to generate a PCR product or the generated product could not be cloned into the T-tailed vector; up to three cloning attempts were made for each GenePair. Supplementary Table 2 gives the complete list of GenePairs and RNAi phenotype class, and indicates whether a clone is available.

Screening using RNAi by feeding

We carried out RNAi as described²⁷. Embryonic lethality was defined as >10% dead embryos, and sterility required a brood size of <10 among fed worms (Ste) or their progeny (Stp); wild-type worms under similar conditions typically have >100 progeny. Each post-embryonic phenotype was required to be present among at least 10% of analysed worms; the phenotypes assayed were Emb (embryonic lethal), Ste (sterile), Stp (sterile progeny), Gro (slow post-embryonic growth), Lva (larval arrest), Lvl (larval lethality), Adl (adult lethal), Bli (blistering of cuticle), Bmd (body morphological defects), Clr (clear), Dpy (dumpy), Egl (egg-laying defective), Him (high incidence of males), Lon (long), Mlt (moult defects), Muv (multivulva), Prz (paralysed), Pvl (protruding vulva), Rol (roller), Rup (ruptured), Sck (sick) and Unc (uncoordinated). Phenotypes expressed in adults (such as Egl) were difficult to score in this screen because food became limiting at this time point; some of the late expressing phenotypes will therefore have been missed. Detailed listings of GenePairs with corresponding RNAi phenotypes are given in Supplementary Tables 3 and 4 and are available at WormBase (<http://www.wormbase.org>).

Bioinformatic analyses

We carried out BlastP³⁵ analyses for all *C. elegans* predicted genes against similar databases (downloaded on 13 Feb 2002) for *S. cerevisiae* (6,183 entries), *Arabidopsis* (25,813 entries), *Drosophila* (13,957 entries) and *Homo sapiens* (36,493 entries), or against *C. elegans* itself. *C. elegans* genes with orthologues were defined as those with BlastP *E* values of less than 10^{−10} with conservation extending over at least 80% of matched protein lengths; 21% of predicted genes in *C. elegans* have such conservation. Predicted gene products were placed into functional classes by manual inspection, primarily using data from Proteome, InterPro release 4.0 (ref. 12) and BLAST analysis^{35,36}. We could place 41% of all predicted genes into 1 of 16 functional classes (Supplementary Table 2), with the remaining 59% having unknown function.

Predicted genes targeted by a given bacterial clone were determined by comparing electronic PCR (ePCR) products corresponding to the bacterial clone insert (<ftp://ftp.ncbi.nlm.nih.gov/pub/schuler/e-PCR>)³⁷ obtained using chromosome DNA files from the W61 release of Wormbase (<ftp://ftp.sanger.ac.uk/pub/wormbase>) to gene predictions from the same database. Roughly 94% of bacterial strains tested correspond to a single predicted gene. To identify genes elsewhere in the genome that might be targeted by cross-

RNAi owing to strong homology of part of the gene to the ePCR product, we found genes having >80% identity over a region of at least 200 nucleotides for each ePCR product by parsing BlastN results against Wormpep release 71. In total, 1,528 clones with RNAi phenotypes could be assigned directly to a single *C. elegans* predicted gene; these are listed in Supplementary Table 3. By contrast, 194 clones with RNAi phenotypes could not be assigned definitively to a single predicted gene; these are listed in Supplementary Table 4 and include GenePairs with either no or multiple ePCR products or for which the ePCR product is not predicted to overlap any coding sequence.

We found chromosomal regions of significant over- or underrepresentation by considering moving windows of 250 consecutive genes along the chromosomes, and by examining whether the number of genes showing a particular phenotype or in a particular expression cluster within a window was significantly different from that expected according to the genomic mean, using a 1% significance level in a two-tailed test using the binomial distribution. Figure 4b and Supplementary Fig. 3 show continuous significant windows, from the midpoint of the leftmost to the midpoint of the rightmost window. Gene positions were taken from the predicted gene set from Wormbase release WS61.

Received 2 July; accepted 29 October 2002; doi:10.1038/nature01278.

1. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
2. Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
3. Maeda, I., Kohara, Y., Yamamoto, M. & Sugimoto, A. Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Curr Biol* **11**, 171–176 (2001).
4. Gonczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).
5. Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).
6. Timmons, L., Court, D. L. & Fire, A. Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263**, 103–112 (2001).
7. Kamath, R. K., Martinez-Campos, M., Zipperlen, P., Fraser, A. G. & Ahringer, J. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *C. elegans*. *Genome Biol.* **2**, 1–10 (2001).
8. Pryer, N. K., Salama, N. R., Schekman, R. & Kaiser, C. A. Cytosolic Sec13p complex is required for vesicle formation from the endoplasmic reticulum *in vitro*. *J. Cell Biol.* **120**, 865–875 (1993).
9. Tavernarakis, N., Wang, S. L., Dorovkov, M., Ryazanov, A. & Driscoll, M. Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. *Nature Genet.* **24**, 180–183 (2000).
10. Piano, F., Schetter, A. J., Mangone, M., Stein, L. & Kempthorne, K. J. RNAi analysis of genes expressed in the ovary of *Caenorhabditis elegans*. *Curr. Biol.* **10**, 1619–1622 (2000).
11. Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **28**, 37–40 (2000).
12. Apweiler, R. *et al.* The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**, 37–40 (2001).
13. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
14. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
15. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
16. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 563–567 (1996) 546.
17. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
18. The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
19. Kuwabara, P. E. Developmental genetics of *Caenorhabditis elegans* sex determination. *Curr. Top. Dev. Biol.* **41**, 99–132 (1999).
20. Reinke, V. *et al.* A global profile of germline gene expression in *C. elegans*. *Mol. Cell* **6**, 605–616 (2000).
21. Kelly, W. G. *et al.* X-chromosome silencing in the germline of *C. elegans*. *Development* **129**, 479–492 (2002).
22. Barnes, T. M., Kohara, Y., Coulson, A. & Hekimi, S. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genetics* **141**, 159–179 (1995).
23. Kim, S. K. *et al.* A gene expression map for *Caenorhabditis elegans*. *Science* **293**, 2087–2092 (2001).
24. Lercher, M. J., Urrutia, A. O. & Hurst, L. D. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature Genet.* **31**, 180–183 (2002).
25. Spellman, P. T. & Rubin, G. M. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.* **1**(1), paper no. 5 (<http://jbiol.com/content/1/1/5>) (2002).
26. Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. M. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature Genet.* **26**, 183–186 (2000).
27. Roy, P. J., Stuart, J., Lund, J. & Kim, S. K. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* **418**, 975–9 (2002).
28. Kruglyak, S. & Tang, H. Regulation of adjacent yeast genes. *Trends Genet.* **16**, 109–111 (2000).
29. Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
30. Donze, O. & Picard, D. RNA interference in mammalian cells using siRNAs synthesized with T7 RNA polymerase. *Nucleic Acids Res.* **30**, e46 (<http://nar.oupjournals.org/cgi/content/full/30/10/e46>) (2002).
31. Elbashir, S. M., Harborth, J., Weber, K. & Tuschl, T. Analysis of gene function in somatic mammalian cells using small interfering RNAs. *Methods* **26**, 199–213 (2002).
32. Paul, C. P., Good, P. D., Winer, I. & Engelke, D. R. Effective expression of small interfering RNA in human cells. *Nature Biotechnol.* **20**, 505–8 (2002).

33. Miyagishi, M. & Taira, K. U6 promoter driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammalian cells. *Nature Biotechnol.* **20**, 497–500 (2002).
34. Sui, G. *et al.* A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc. Natl Acad. Sci. USA* **99**, 5515–5520 (2002).
35. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
36. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
37. Schuler, G. D. Sequence mapping by electronic PCR. *Genome Res.* **7**, 541–550 (1997).
38. Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215 (2000).
39. Wood, V. *et al.* The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**, 871–880 (2002).

Supplementary Information accompanies the paper on *Nature's* website
(<http://www.nature.com/nature>).

Acknowledgements We thank T. Blumenthal, B. Kelly, S. Kim and V. Reinke for discussions and sharing data before publication; B. Shumacher for assistance early in the project; and G. Chalkin, J. Peacock, J. Abbott and K. Rossiter for preparing the media. R.S.K. was supported by a Howard Hughes Medical Institute Predoctoral Fellowship; A.G.F. by a US Army Breast Cancer Research Fellowship; Y.D., R.D., M.G., D.P.W. and P.Z. by the Wellcome Trust; G.P. by the Canadian Institute of Health Research and the Wellcome Trust; A.K. by the European Molecular Biology Laboratory; N.L.B. by the European Molecular Biology Organization; S.M. by the Centro de Investigacion del Cancer; M.S. by a Swiss National Science Foundation fellowship and J.A. by a Wellcome Trust Senior Research Fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.
(e-mail: jaa@mole.bio.cam.ac.uk).

Quantum key distribution using gaussian-modulated coherent states

Frédéric Grosshans*, Gilles Van Assche†, Jérôme Wenger*, Rosa Brouni*, Nicolas J. Cerf† & Philippe Grangier*

* Laboratoire Charles Fabry de l'Institut d'Optique, CNRS UMR 8501, 91403 Orsay, France

† Ecole Polytechnique, CP 165, Université Libre de Bruxelles, 1050 Bruxelles, Belgium

Quantum continuous variables¹ are being explored^{2–14} as an alternative means to implement quantum key distribution, which is usually based on single photon counting¹⁵. The former approach is potentially advantageous because it should enable higher key distribution rates. Here we propose and experimentally demonstrate a quantum key distribution protocol based on the transmission of gaussian-modulated coherent states (consisting of laser pulses containing a few hundred photons) and shot-noise-limited homodyne detection; squeezed or entangled beams are not required¹³. Complete secret key extraction is achieved using a reverse reconciliation¹⁴ technique followed by privacy amplification. The reverse reconciliation technique is in principle secure for any value of the line transmission, against gaussian individual attacks based on entanglement and quantum memories. Our table-top experiment yields a net key transmission rate of about 1.7 megabits per second for a loss-free line, and 75 kilobits per second for a line with losses of 3.1 dB. We anticipate that the scheme should remain effective for lines with higher losses, particularly because the present limitations are essentially technical, so that significant margin for improvement is available on both the hardware and software.

Much interest has arisen recently in using the electromagnetic field amplitudes to obtain possibly more efficient quantum continuous variable (QCV) alternatives^{2–14} to the usual photon-counting quantum key distribution (QKD) techniques (see ref. 15 and references therein)—for instance, by using ‘non-classical’ light

beams^{2–11}. In fact, it was shown in ref. 13 that squeezed or entangled light is not required to achieve this goal: an equivalent level of security may be obtained by transmitting ‘quasi-classical’ coherent states. When the line transmission is larger than 50% (line loss ≤ 3 dB), the physical limits on QCV cloning^{16–18} ensure that this protocol is secure against individual attacks. This corroborates the fact that QKD only requires non-orthogonal states, and may well work with macroscopic signals instead of single photons¹⁹. There are in principle various ways for the partners Alice and Bob to distribute keys beyond this 3 dB limit, for instance by using entanglement purification²⁰ or postselection¹². Therefore these QCV schemes stimulate many fundamental questions about the physical origin of QKD security. As will be shown below, cryptographic security appears to have a strong relationship with entanglement, even though our protocol does not rely on entangled states.

Here we introduce and implement a coherent-state QKD protocol, and we demonstrate that it is, in principle, secure for any value of the line transmission. It relies on the distribution of a gaussian key⁷ obtained by continuously modulating the phase and amplitude of coherent light pulses¹³ at Alice's side, and subsequently performing homodyne detection at Bob's side. The continuous data are then converted into a common binary key via a specifically designed reconciliation algorithm^{8,10}. The security against arbitrarily high losses is achieved by reversing the reconciliation protocol, that is, Alice attempts to guess what was received by Bob rather than Bob guessing what was sent by Alice. Such a reverse reconciliation protocol¹⁴ gives Alice an advantage over a potential eavesdropper Eve, regardless of the line loss. The practical limitations of our scheme are essentially technical, and appear to be due mostly to the current efficiency of the reconciliation software.

The protocol runs as follows¹³. First, Alice draws two random numbers x_A and p_A from a gaussian distribution of mean zero and variance $V_A N_0$, where N_0 denotes the shot-noise variance. Then, she sends the coherent state $|x_A + ip_A\rangle$ to Bob, who randomly chooses to measure either quadrature x or p . Later, using a public authenticated channel, he informs Alice about which quadrature he measured, so she may discard the irrelevant data. After many similar exchanges, Alice and Bob (and possibly the eavesdropper Eve) share a set of correlated gaussian variables, which we call ‘key elements’.

Classical data processing is then necessary for Alice and Bob to obtain a fully secret binary key. First, Alice and Bob publicly

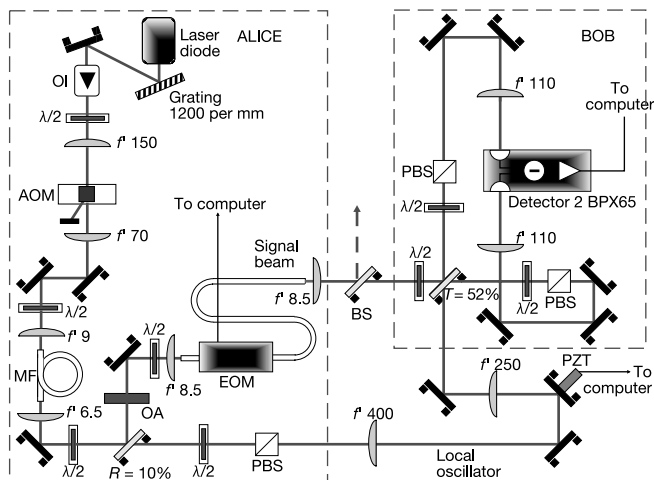


Figure 1 Experimental set-up. Laser diode, SDL 5412 (780 nm); OI, optical isolator; $\lambda/2$, half-wave plate; AOM, acousto-optic modulator; MF, polarization maintaining single-mode fibre; OA, optical attenuator; EOM, electro-optic amplitude modulator; PBS, polarizer; BS, beam splitter; PZT, piezoelectric transducer. Focal lengths (f') are given in millimetres. R and T are reflection and transmission coefficients.

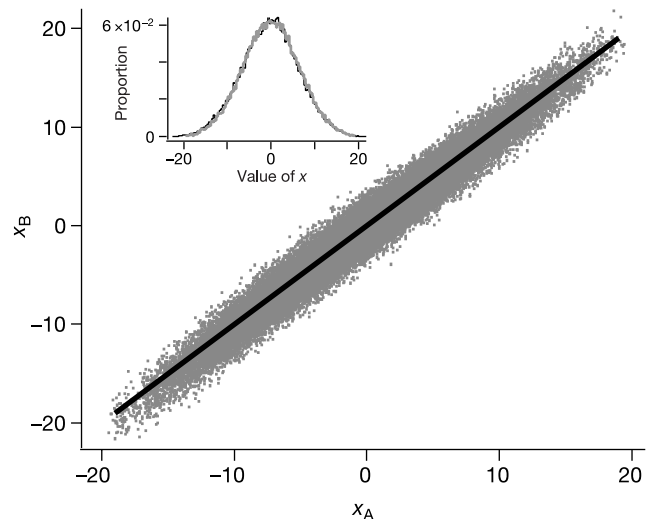


Figure 2 Bob's measured quadrature as a function of the amplitude sent by Alice (in Bob's measurement basis) for a burst of 60,000 pulses. The line transmission is 100% and the modulation variance is $V = 41.7$. The solid line represents the expected unity slope. Inset, the corresponding histograms of Alice's (grey curve) and Bob's (black curve) data.

compare a random sample of their key elements to evaluate the error rate and transmission efficiency of the quantum channel. From the observed correlations, Alice and Bob evaluate the amount of information they share ($I_{AB} = I_{BA}$) and the maximum information Eve may have obtained (by eavesdropping) about their values (I_{AE} and I_{BE}). It is known that Alice and Bob can, in principle, distil from their data a common secret key of size $S > \sup(I_{AB} - I_{AE}, I_{BA} - I_{BE})$ bits per key element^{21,22}. This requires classical communication over an authenticated public channel, and may be divided into two steps: reconciliation (that is, correcting the errors while minimizing the information revealed to Eve) and privacy amplification (that is, making the key secret). As we deal here with continuous data, we developed a ‘sliced’ reconciliation algorithm^{8,10} to extract common bit strings from the correlated key elements. In order to reconcile Bob’s measured data with Alice’s sent data, the most natural way to proceed is that Bob gets R extra bits of information from Alice in order to correct the transmission errors. The corresponding direct reconciliation (DR) protocols, which have been used so far in QCV QKD^{7,13}, allow the generation of a common string of $I_{AB} + R$ bits, of which Eve may know up to $I_{AE} + R$ bits. Here we rather consider reverse reconciliation (RR) protocols¹⁴, where Bob sends R bits of information to Alice so that she incorporates the transmission errors in her initial data. These RR protocols allow the generation of a common string of $I_{BA} + R$ bits, of which Eve may know $I_{BE} + R$ bits. This turns out to be particularly well suited to QCV QKD, because it is more difficult for Eve to control the errors at Bob’s side than to read Alice’s modulation. The last step of key extraction, namely privacy amplification, consists of filtering out Eve’s information by properly mixing the reconciled bits to spread Eve’s uncertainty over the entire final key. This procedure requires an estimate of Eve’s information on the reconciled key, so we need a bound on I_{AE} for DR, or I_{BE} for RR. In addition, Alice and Bob must keep track of the information publicly revealed during reconciliation. This knowledge is destroyed at the end of the privacy amplification procedure, reducing the key length by the same amount. The DR bound¹³ on I_{AE} implies that the security cannot be warranted if the line transmission G is below 50%. We will now establish the RR bound on I_{BE} , and show that it is not associated with a minimum value of G .

In a RR scheme, Eve needs to guess Bob’s measurement outcome without adding too much noise on his data. This can be done via an ‘entangling cloner’, which creates two quantum-correlated copies of Alice’s quantum state, so Eve simply keeps one of them while sending the other to Bob. Let (x_{in}, p_{in}) be the input field quadratures of the entangling cloner, and (x_B, p_B) , (x_E, p_E) the quadratures of Bob’s and Eve’s output fields. To be safe, we must assume Eve uses the best possible entangling cloner compatible with the

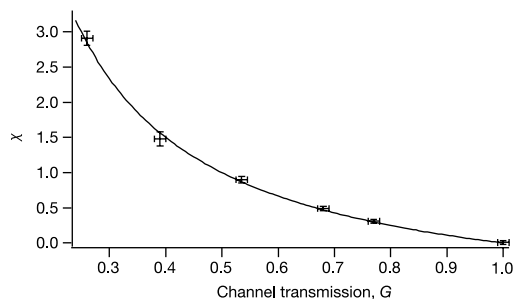


Figure 3 Channel equivalent noise χ_{line} as a function of line transmission G . The curve is the theoretical prediction $\chi_{vac} = (1 - G)/G$. The error bars include two contributions with approximately the same size, from statistics (evaluated over blocks of 60,000 pulses) and systematics (calibration errors and drifts).

parameters of the Alice–Bob channel: Eve’s cloner should minimize the conditional variances^{23,24} $V(x_B|x_E)$ and $V(p_B|p_E)$, that is, the variances of Eve’s estimates of Bob’s field quadratures (x_B, p_B). These variances are constrained by Heisenberg-type relations (see Methods), which limit what can be obtained by Eve:

$$V(x_B|x_A) V(p_B|p_E) \geq N_0^2 \quad \text{and} \quad V(p_B|p_A) V(x_B|x_E) \geq N_0^2 \quad (1)$$

where $V(x_B|x_A)$ and $V(p_B|p_A)$ denote Alice’s conditional variances. This means that Alice and Eve cannot jointly know more about Bob’s conjugate quadratures than is allowed by the uncertainty principle. Now, Alice’s variances can be bounded by using the measured parameters of the quantum channel, which in turn makes it possible to bound Eve’s variances.

The channel is described by the linear relations $x_B = G_x^{1/2}(x_{in} + B_x)$ and $p_B = G_p^{1/2}(p_{in} + B_p)$, with $\langle x_{in}^2 \rangle = \langle p_{in}^2 \rangle = V N_0 \geq N_0$, $\langle B_{x,p}^2 \rangle = \chi_{x,p} N_0$, and $\langle x_{in} B_x \rangle = \langle p_{in} B_p \rangle = 0$. Here χ_x, χ_p represent the channel noises referred to its input, called equivalent input noises^{23,24}, while G_x, G_p are the channel gains in x and p , and V is the variance of Alice’s field quadratures in shot-noise units ($V = V_A + 1$). The output–output correlations of the entangling cloner, described by $V(x_B|x_E)$ and $V(p_B|p_E)$, depend only on the density matrix D_{in} of the input field (x_{in}, p_{in}), and not on the way it is produced, namely whether it is a gaussian mixture of coherent states or one of two entangled beams. Inequalities (1) thus have to be fulfilled for all physically allowed values of $V(x_B|x_A)$ and $V(p_B|p_A)$, given D_{in} . Therefore, the values of $V(x_A|x_B)$ and $V(p_A|p_B)$ that should be used in inequalities (1) to limit Eve’s knowledge are the minimum values Alice might achieve by using the maximal entanglement compatible with V , namely (see Methods):

$$V(x_B|x_A)_{min} = G_x(\chi_x + V^{-1})N_0 \quad (2)$$

$$V(p_B|p_A)_{min} = G_p(\chi_p + V^{-1})N_0$$

These lower bounds are thus directly connected with entanglement, even though Alice does not use it in practice. They may be compared with the actual values when Alice sends coherent states, that is, $V(x_B|x_A)_{coh} = G_x(\chi_x + 1)N_0$ and $V(p_B|p_A)_{coh} = G_p(\chi_p + 1)N_0$. The lower bounds on Eve’s conditional variances are then obtained from equations (1) and (2), as:

$$V(p_B|p_E) \geq N_0 / \{G_x(\chi_x + V^{-1})\} \quad (3)$$

$$V(x_B|x_E) \geq N_0 / \{G_p(\chi_p + V^{-1})\}$$

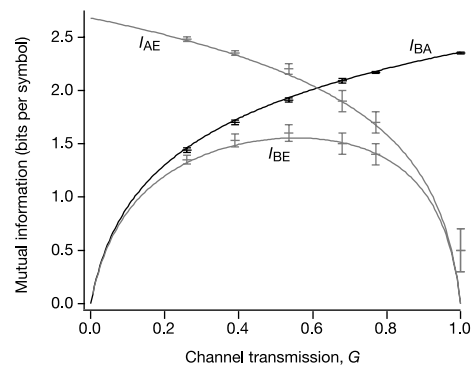


Figure 4 Values of I_{BA} , I_{BE} and I_{AE} as a function of the line transmission G for $V \approx 40$. Here, I_{BA} is given by equation (4a), including all transmission and detection noises for evaluating V_B and $(V_{B|A})_{coh}$. The expression for I_{BE} is given by equation (4b), using the same V_B and $(V_{B|E})_{min} = N_0 / \{G(\chi_{line} + V^{-1})\} + N_{el} + N_{hom}$. This expression realistically assumes that Eve cannot know the noises N_{el} and N_{hom} , which are internal to Bob’s detection set-up. For comparison with DR, the value of I_{AE} is also plotted (the theoretical value of I_{AE} is obtained from ref. 13).

Table 1 **Ideal and practical net secret key rates**

V	G _{line}	Losses (dB)	I _{BA} (bit)	I _{BE} (% I _{BA})	I _{rec} (% I _{BA})	Ideal RR rate (kbit s ⁻¹)	Practical RR rate (kbit s ⁻¹)	Ideal DR rate (kbit s ⁻¹)	Practical DR rate (kbit s ⁻¹)
41.7	1	0	2.39	0	88	1,920	1,690	1,910	1,660
38.6	0.79	1.0	2.17	58	85	730	470	540	270
32.3	0.68	1.7	1.93	67	79	510	185	190	—
27	0.49	3.1	1.66	72	78	370	75	0	—
43.7	0.26	5.9	1.48	93	71	85	—	0	—

The parameters of the quantum key exchange are measured for several values of the channel transmission G (the corresponding losses are also given in decibels). The variations of the variance V of Alice's field quadrature are due to different experimental adjustments. The information I_{BA} is given in bits per time slot. Also shown are the maximum information gained by Eve (I_{BE}) and the extracted information by reverse reconciliation (I_{rec}). The ideal secret key bit rates would be obtained from our measured data with perfect key distillation that yields exactly $I_{AB} - I_{BE}$ bits (RR) or $I_{AB} - I_{AE}$ bits (DR), whereas the practical secret key bit rates are the one achieved with our current key distillation procedure ('—' means that no secret key is generated). Both bit rates are calculated over bursts of about 60,000 pulses at 800 kHz, not taking into account the duty cycle (~5%) in the present set-up.

A physical realization of an entangling cloner reaching these bounds is sketched in ref. 14.

To assess the security of the RR scheme, we assume that Eve's ability to infer Bob's measurement can reach the limit put by inequalities (3). For simplicity, we consider the channel gains and noises and the signal variances to be the same for x and p (in practice, deviations should be estimated by statistical tests). The information rates can be derived using Shannon's theory for gaussian additive-noise channels²⁵, giving

$$I_{BA} = (1/2) \log_2 [V_B / (V_{B|A})_{\text{coh}}] \\ = (1/2) \log_2 [(V + \chi) / (1 + \chi)] \quad (4a)$$

$$I_{BE} = (1/2) \log_2 [V_B / (V_{B|E})_{\text{min}}] \\ = (1/2) \log_2 [G^2 (V + \chi) (V^{-1} + \chi)] \quad (4b)$$

expressed in bits per symbol (or per key element). Here $V_B = \langle x_B^2 \rangle = \langle p_B^2 \rangle = G(V + \chi)N_0$ is Bob's variance, $(V_{B|E})_{\text{min}} = V(x_B|x_E)_{\text{min}} = V(p_B|p_E)_{\text{min}} = N_0 / \{G(\chi + V^{-1})\}$ is Eve's minimum conditional variance, and $(V_{B|A})_{\text{coh}} = V(x_B|x_A)_{\text{coh}} = V(p_B|p_A)_{\text{coh}} = G(\chi + 1)N_0$ is Alice's conditional variance for a coherent-state protocol. The secret bit rate of a RR protocol is thus

$$\Delta I_{RR} = I_{BA} - I_{BE} = -(1/2) \log_2 [G^2 (1 + \chi) (V^{-1} + \chi)] \quad (5)$$

and the security is guaranteed if $\Delta I_{RR} > 0$. The equivalent input noise χ can be split into a 'vacuum noise' component due to the line losses, given by $\chi_{\text{vac}} = (1 - G)/G$, and an 'excess noise' component defined as $\varepsilon = \chi - \chi_{\text{vac}}$. In the high-loss limit ($G \ll 1$), the RR protocol remains secure if $\varepsilon < (V - 1)/(2V) \approx 1/2$, that is, if the amount of excess noise ε is not too large. In contrast, a DR protocol requires low-loss lines, as the security is warranted only if $\chi < 1$, that is, if $G > 1/(2 - \varepsilon)$. Note that DR tolerates an excess noise up to $\varepsilon \approx 1$, so it might be preferred to RR for low-loss but noisy channels.

Our experimental implementation (Fig. 1) of the quantum key exchange uses 120-ns coherent pulses at a 800-kHz repetition rate (wavelength of 780 nm, see Methods). Data bursts of 60,000 pulses have been analysed (Fig. 2). For each burst, a subset of the values are disclosed to evaluate the transmission G and the total added noise variance. The output noise has four contributions: the shot noise N_0 , the channel noise $\chi_{\text{line}}N_0$, the electronics noise of Bob's detector ($N_{\text{el}} = 0.33N_0$), and the noise due to imperfect homodyne detection efficiency ($N_{\text{hom}} = 0.27N_0$). When introducing line losses using a variable attenuator, the measured χ_{line} increases as $(1 - G)/G$, as shown in Fig. 3 ($\varepsilon_{\text{line}} = 0$ here). The two detection noises N_{el} and N_{hom} originate from Bob's detection system, so they must be taken into account when estimating I_{BA} . In contrast, we may reasonably assume that Eve cannot know or control the corresponding fluctuations, so her attack is inferred on the basis of the line noise χ_{line} only (see Supplementary Information for details). Figure 4 shows explicitly the mutual information

between all parties, which makes straightforward the comparison between the DR and RR protocols.

We wrote a computer program that implements the reconciliation algorithm followed by privacy amplification (see Methods and Supplementary Information). Although Alice and Bob are not spatially separated in the present set-up, the analysed data have the same structure as in a realistic cryptographic exchange. Table 1 shows the ideal and practical net key rates of our reverse QKD protocol, as well as the DR values for comparison. The RR scheme is efficient for any value of G provided that the reconciliation protocol achieves the limit given by I_{BA} . However, unavoidable deviations of the algorithm from Shannon's limit reduce the actual reconciled information I_{rec} between Alice and Bob, while I_{BE} is of course assumed unaffected. For high modulation ($V \approx 40$) and low losses, the reconciliation efficiency lies around 80%, which makes it possible to distribute a secret key at a rate of several hundreds of kilobits per second. However, the achievable reconciliation efficiency drops when the signal-to-noise ratio decreases, but this can be improved by reducing the modulation variance, which increases the ratio I_{BA}/I_{BE} . Although the ideal secret key rate is then lower, we could process the data with a reconciliation efficiency of 78% for $G = 0.49$ (3.1 dB) and $V = 27$, resulting in a net key rate of 75 kbit s⁻¹ (see also Methods). This clearly demonstrates that RR continuous-variable protocols operate efficiently at and beyond the 3 dB loss limit of DR protocols. We emphasize that this result is obtained despite the fact that the evaluated reconciliation cost is higher for RR than for DR: the better result for RR is essentially due to its initial 'quantum advantage'.

In photon-counting QKD, the key rate is limited by the single-photon detectors, in which the avalanche processes are difficult to control reliably at very high counting rates. In contrast, homodyne detection may run at frequencies up to tens of MHz. In addition, a specific advantage of the high dimensionality of the QCV phase space is that the field quadratures can be modulated with a large dynamics, allowing the encoding of several key bits per pulse (see Table 1). Very high secret bit rates are therefore attainable with our coherent-state protocol on low-loss lines. For high-loss lines, our protocol is at present limited by the reconciliation efficiency, but its intrinsic performances remain very high. Because most of the limitations of the present proof-of-principle experiment appear to be of a technical nature, there is still a considerable margin for improvement, both in the hardware (increased detection bandwidth, better homodyne efficiency, lower electronic noise), and in the software (better reconciliation algorithms²⁶, see Methods). In conclusion, the way seems open for implementing the present proposal at telecommunications wavelengths as a practical, high-bit-rate, quantum key distribution scheme over long distances. □

Methods

Relevant Heisenberg relations

In a RR protocol, Alice's estimator for x_B and Eve's estimator for p_B can be denoted respectively as αx_A and βp_E , where α, β are real numbers. The corresponding errors are

$x_{B|A,\alpha} = x_B - \alpha x_A$, and $p_{B|E,\beta} = p_B - \beta p_E$. Because Alice's, Bob's and Eve's operators commute, we have $[x_{B|A,\alpha}, p_{B|E,\beta}] = [x_B, p_B]$, and thus the Heisenberg relation $\Delta x_{B|A,\alpha}^2 \Delta p_{B|E,\beta}^2 \geq N_0^2$. Defining the conditional variances as $V(x_B|x_A) = \min_{\alpha} \{\Delta x_{B|A,\alpha}^2\}$ and $V(p_B|p_E) = \min_{\beta} \{\Delta p_{B|E,\beta}^2\}$, we obtain $V(x_B|x_A) V(p_B|p_E) \geq N_0^2$, or, by exchanging x and p , $V(p_B|p_A) V(x_B|x_E) \geq N_0^2$.

Alice has the estimators (x_A, p_A) for the field $(x_{in}, p_{in}) = (x_A + A_x, p_A + A_p)$ that she sends, with $\langle A_x^2 \rangle = \langle A_p^2 \rangle = s N_0$. Here s measures the amount of squeezing possibly used by Alice in her state preparation¹⁴, with $s \geq V^{-1}$ for consistency with Heisenberg's relations. By calculating $\langle p_A^2 \rangle = (V - s) N_0$, $\langle p_B^2 \rangle = G_p (V + \chi_p) N_0$, $\langle p_A p_B \rangle = G_p^{1/2} \langle p_A^2 \rangle$, we obtain the conditional variance $V(p_B|p_A) = \langle p_B^2 \rangle - \langle p_A p_B \rangle^2 / \langle p_A^2 \rangle = G_p (s + \chi_p) N_0$. This equation and the constraint $s \geq V^{-1}$ gives $V(p_B|p_A) \geq G_p (V^{-1} + \chi_p) N_0$, and similarly $V(x_B|x_A) \geq G_x (V^{-1} + \chi_x) N_0$. The bound on $V_{B|A}$ is thus obtained by assuming that Alice may use squeezed or entangled beams, while the bound on $V_{B|E}$ can only be achieved if Eve uses an entangling attack. This reflects the fact that squeezing or entanglement play a crucial role in our security demonstration, even though the protocol implies coherent states. Our security proof addresses individual gaussian attacks only, but as the entangling cloner attack saturates the Heisenberg uncertainty relations, we conjecture that it encompasses all incoherent (non-collective) eavesdropping strategies.

Experimental set-up

A continuous-wave laser diode at 780 nm wavelength associated with an acousto-optic modulator is used to emit 120-ns (full-width at half-maximum) pulses at a 800 kHz rate. The signal pulses contain up to 250 photons, while the local oscillator (LO) power is 1.3×10^8 photons per pulse. The amplitude of each pulse is arbitrarily modulated by an integrated electro-optic modulator. However, owing to the unavailability of a fast phase modulator at 780 nm, the phase is not randomly modulated but scanned continuously. No genuine secret key can be distributed, strictly speaking, but random permutations of Bob's data are used to provide realistic data (see Supplementary Information). The data are organized in bursts of 60,000 pulses, separated by synchronization periods also used to lock the phase of the LO. The overall homodyne detection efficiency is 0.81, due to the optical transmission (0.92), the mode-matching efficiency (0.96) and the photodiode quantum efficiency (0.92). For the critical data at 3 dB loss, the mode-matching efficiency was improved to 0.99, and thus the overall efficiency was 0.84. We also point out that many blocks of data were exchanged around the 3 dB loss point, with a typical rate above 55 kbit s^{-1} .

Secret key distillation

A common bit string is extracted from the continuous data by sequentially reconciling several strings ('slices') of binary functions of the gaussian key elements, applying a binary reconciliation protocol successively on each bit^{8,10}. Here, we used five slices, each being corrected either by a trivial one-way protocol (communicating the bits) when the bit error rate (BER) is high, or by the two-way protocol Cascade^{27,28} when the BER is low. Note that the disclosed slices are useful for reconciling the remaining slices with less information leaking to Eve, even though they of course do not yield secret bits as such. In addition, Alice and Bob encrypt their classical messages using the one-time pad scheme with a fraction of the previous key bits, or a bootstrap key for the first block. For slices corrected with Cascade, the exchanged parities are encrypted with the same key bits on both sides²⁹, making Eve aware of the differences between Alice's and Bob's parities (that is, the error positions) but not of their individual values. Fully communicated slices are also encrypted, thereby revealing no information at all to Eve. Still, Eve may exploit the interactivity of Cascade and gain some information on the final key by combining her knowledge of the error positions with that of the correlations between Alice's and Bob's gaussian values. In the present protocol, this information is numerically calculated for an entangling cloner attack, and then destroyed by privacy amplification. This is achieved by appropriate 'hashing'³⁰ functions (see Supplementary Information). The resulting net secret key rate is then obtained by subtracting, from the raw key rate, the cost of the one-time pad encryption and the error-position information. Finally, we emphasize that sliced reconciliation can be made very close to a one-way protocol by increasing the number of key elements from which the bits are jointly extracted (multidimensional reconciliation⁸). This approach was not implemented here, but should deliver an improved secret key rate, approaching the value from the Csizár-Körner formula^{21,22}.

Received 8 July; accepted 30 October 2002; doi:10.1038/nature01289.

1. Braunstein, S. L. & Pati, A. K. *Quantum Information Theory with Continuous Variables* (Kluwer Academic, Dordrecht, in the press).
2. Hillery, M. Quantum cryptography with squeezed states. *Phys. Rev. A* **61**, 022309 (2000).
3. Ralph, T. C. Continuous variable quantum cryptography. *Phys. Rev. A* **61**, 010303(R) (2000).
4. Ralph, T. C. Security of continuous-variable quantum cryptography. *Phys. Rev. A* **62**, 062306 (2000).
5. Reid, M. D. Quantum cryptography with a predetermined key, using continuous-variable Einstein-Podolsky-Rosen correlations. *Phys. Rev. A* **62**, 062308 (2000).
6. Gottesman, D. & Preskill, J. Secure quantum key distribution using squeezed states. *Phys. Rev. A* **63**, 022309 (2001).
7. Cerf, N. J., Lévy, M. & Van Assche, G. Quantum distribution of Gaussian keys using squeezed states. *Phys. Rev. A* **63**, 052311 (2001).
8. Van Assche, G., Cardinal, J. & Cerf, N. J. Reconciliation of a quantum-distributed Gaussian key. Preprint cs.CR/0107030 at (<http://arxiv.org>) (2001).
9. Bencheikh, K., Symul, Th., Jankovic, A. & Levenson, J. A. Quantum key distribution with continuous variables. *J. Mod. Opt.* **48**, 1903–1920 (2001).
10. Cerf, N. J., Iblisdir, S. & Van Assche, G. Cloning and cryptography with quantum continuous variables. *Eur. Phys. J. D* **18**, 211–218 (2002).
11. Silberhorn, Ch., Korolkova, N. & Leuchs, G. Quantum key distribution with bright entangled beams. *Phys. Rev. Lett.* **88**, 167902 (2002).

12. Silberhorn, Ch., Ralph, T. C., Lütkenhaus, N. & Leuchs, G. Continuous variable quantum cryptography beating the 3 dB loss limit. *Phys. Rev. Lett.* **89**, 167901 (2002).
13. Grosshans, F. & Grangier, Ph. Continuous variable quantum cryptography using coherent states. *Phys. Rev. Lett.* **88**, 057902 (2002).
14. Grosshans, F. & Grangier, Ph. Reverse reconciliation protocols for quantum cryptography with continuous variables. Preprint quant-ph/0204127 at (<http://arxiv.org>) (2002).
15. Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74**, 145–195 (2002).
16. Cerf, N. J., Ipe, A. & Rottenberg, X. Cloning of continuous variables. *Phys. Rev. Lett.* **85**, 1754–1757 (2000).
17. Cerf, N. J. & Iblisdir, S. Optimal N-to-M cloning of conjugate quantum variables. *Phys. Rev. A* **62**, 040301(R) (2000).
18. Grosshans, F. & Grangier, Ph. Quantum cloning and teleportation criteria for continuous quantum variables. *Phys. Rev. A* **64**, 010301(R) (2001).
19. Bennett, C.-H. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett.* **68**, 3121–3124 (1992).
20. Duan, L.-M., Giedke, G., Cirac, J. I. & Zoller, P. Entanglement purification of gaussian continuous variable quantum states. *Phys. Rev. Lett.* **84**, 4002–4005 (2000).
21. Csizár, I. & Körner, J. Broadcast channel with confidential messages. *IEEE Trans. Inform. Theory* **24**, 339–348 (1978).
22. Maurer, U. M. Secret key agreement by public discussion from common information. *IEEE Trans. Inform. Theory* **39**, 733–742 (1993).
23. Poizat, J.-Ph., Roch, J.-F. & Grangier, Ph. Characterization of quantum non-demolition measurements in optics. *Ann. Phys. (Paris)* **19**, 265–297 (1994).
24. Grangier, Ph., Levenson, J. A. & Poizat, J.-Ph. Quantum non-demolition measurements in optics. *Nature* **396**, 537–542 (1998).
25. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 623–656 (1948).
26. Buttler, W. T., Lamoreaux, S. K., Torgerson, J. R., Nickel, G. H. & Peterson, C. G. Fast, efficient error reconciliation for quantum cryptography. Preprint quant-ph/0203096 at (<http://arxiv.org>) (2002).
27. Brassard, G. & Salvail, L. *Advances in Cryptology – Eurocrypt '93 Lecture Notes in Computer Science* (ed. Hellese, T.) 411–423 (Springer, New York, 1993).
28. Nguyen, K. *Extension des Protocoles de Réconciliation en Cryptographie Quantique* Thesis, Univ. Libre de Bruxelles (2002).
29. Lo, H.-K. Method for decoupling error correction from privacy amplification. Preprint quant-ph/0201030 at (<http://arxiv.org>) (2002).
30. Bennett, C. H., Brassard, G., Crépeau, C. & Maurer, U. M. Generalized privacy amplification. *IEEE Trans. Inform. Theory* **41**, 1915–1935 (1995).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements The contributions of J. Gao to the early stages of the experiment, and of K. Nguyen to the software development, are acknowledged. We thank S. Iblisdir for discussions, and Th. Debuisschert for the loan of the 780 nm integrated modulator. This work was supported by the EU programme IST/FET/QIPC (projects "QUICOV" and "EQUIP"), the French programmes ACI Photonique and ASTRE, and by the Belgian programme ARC.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.G. (e-mail: philippe.grangier@iota.u-psud.fr).

Single-nanowire electrically driven lasers

Xiangfeng Duan*†, Yu Huang*†, Ritesh Agarwal* & Charles M. Lieber*‡

* Department of Chemistry and Chemical Biology, ‡ Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA
† These authors contributed equally to this work

Electrically driven semiconductor lasers are used in technologies ranging from telecommunications and information storage to medical diagnostics and therapeutics¹. The success of this class of lasers is due in part to well-developed planar semiconductor growth and processing, which enables reproducible fabrication of integrated, electrically driven devices^{2,3}. Yet this approach to device fabrication is also costly and difficult to integrate directly with other technologies such as silicon microelectronics. To overcome these issues for future applications, there has been considerable interest in using organic molecules^{4,5}, polymers^{6,7},

and inorganic nanostructures^{8–10} for lasers, because these materials can be fashioned into devices by chemical processing. Indeed, amplified stimulated emission and lasing have been reported for optically pumped organic systems^{4–7} and, more recently, inorganic nanocrystals^{8,9} and nanowires¹⁰. However, electrically driven lasing, which is required in most applications, has met with several difficulties in organic systems¹¹, and has not been addressed for assembled nanocrystals or nanowires. Here we investigate the feasibility of achieving electrically driven lasing from individual nanowires. Optical and electrical measurements made on single-crystal cadmium sulphide nanowires show that these structures can function as Fabry–Perot optical cavities with mode spacing inversely related to the nanowire length. Investigations of optical and electrical pumping further indicate a threshold for lasing as characterized by optical modes with instrument-limited linewidths. Electrically driven nanowire lasers, which might be assembled in arrays capable of emitting a wide range of colours, could improve existing applications and suggest new opportunities.

Free-standing semiconductor nanowires, which can be prepared as single crystals with controlled diameters by metal nanocluster catalysed growth^{12–14}, are attractive building blocks for creating electrically driven lasers because their defect-free structures exhibit the superior electrical transport of high-quality planar inorganic devices^{15–17}, and because a single nanowire can function as a stand-alone optical cavity and gain medium. Here we focus on single-crystal, 80–200 nm diameter cadmium sulphide (CdS) nanowires that have a wurtzite structure with a [001] growth axis. In general, a nanowire will function as a single-mode optical waveguide¹⁸ (Fig. 1a) when $1 \approx (\pi D/\lambda)(n_1^2 - n_0^2)^{0.5} < 2.4$, where 1 is a practical lower limit, D is the nanowire diameter, λ is the wavelength, and n_1 and n_0 are the refractive indices of the nanowire and surrounding medium, respectively. In the case of CdS nanowires ($n_1 \approx 2.5$, $\lambda \approx 510$ nm, 300 K), the minimum diameter needed to support a single mode is of the order of 70 nm. If the ends of the nanowire are cleaved, they can function as two reflecting mirrors that define a Fabry–Perot optical cavity with modes $m(\lambda/2n_1) = L$, where m is an integer and L is the length of the cavity. Significantly, transmission and scanning electron microscopy studies show that solution-phase sonication of CdS nanowires produces a high (>50%) yield of flat ends (Fig. 1b), indicative of cleavage perpendicular to the [001] growth direction. These results suggest that a substantial number of nanowires should function as Fabry–Perot cavities.

The optical cavity properties of the CdS nanowires, which are central to our use of these nanostructures for lasers, were characterized by photoluminescence measurements at the single nanowire level using a far-field epifluorescence microscope¹⁹. A typical room-temperature luminescence image (Fig. 1c) of a CdS nanowire excited with a tightly focused laser about 15 μm from the nanowire end shows strong emission at the excitation focus and also prominent emission near the nanowire end. Studies of several nanowires show that outside the excitation region emission is only observed from the nanowire ends, thus suggesting that these CdS nanowires function as waveguides.

To further probe the nanowire cavity properties, spectroscopy measurements have been made at different regions as a function of excitation power under uniform illumination. At low power, photoluminescence spectra recorded from the body exhibit a broad peak with a maximum at 512 nm and a full-width at half-maximum (FWHM) of 24 nm (Fig. 1d). The peak maximum is consistent with room-temperature band-edge emission from CdS, and contrasts with the deep-level emission at around 600 nm that usually dominates epitaxial CdS thin films²⁰. Spectra recorded from the nanowire end at low excitation power showed a relatively broad peak that was red-shifted about 30 nm relative to spectra from the body. The observed spectra red-shift is consistent with re-absorption of band-edge emission within the CdS nanowire cavity.

Photoluminescence measurements made at higher excitation powers reveal other important features about the CdS nanowire cavities (Fig. 1e). First, the nanowire end emission blue-shifts towards the band edge as the re-absorption is partially saturated with increasing excitation power. Second, the end-emission intensity increases superlinearly with excitation power, whereas emission from the nanowire body exhibits a slight, approximately linear increase. Third, periodic variations in the intensity, which are suggestive of the longitudinal modes of a Fabry–Perot cavity, are observed. For a cavity of length L , the mode spacing, $\Delta\lambda$, is given by

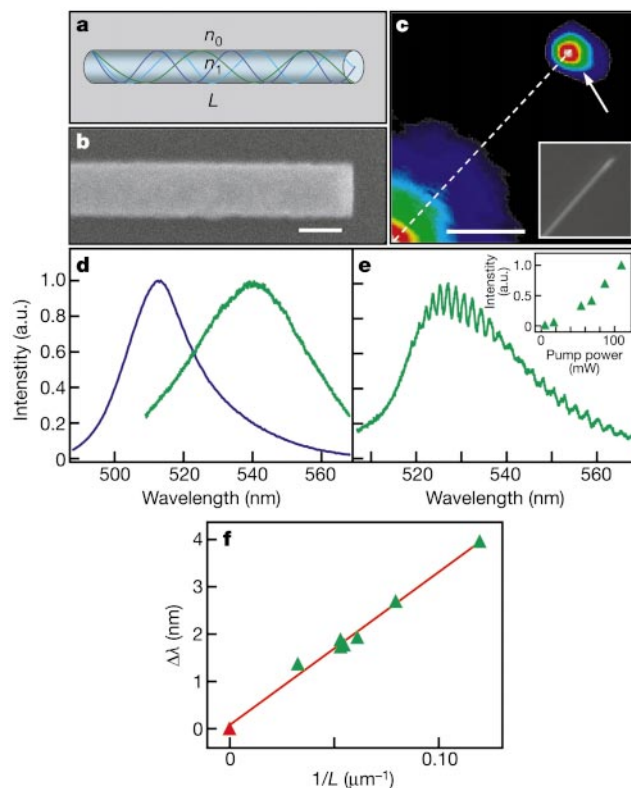


Figure 1 Nanowire Fabry–Perot optical cavities. **a**, Schematic showing a nanowire as an optical waveguide, and with cleaved ends defining a Fabry–Perot cavity. **b**, SEM image of a cleaved CdS nanowire end. Scale bar, 100 nm. **c**, Room-temperature photoluminescence image of a CdS nanowire excited (low-left corner, power 10 mW) about 15 μm away from the nanowire end. The white arrow and dashed line highlight the nanowire end and axis, respectively. Scale bar, 5 μm . Inset, an optical image of the nanowire obtained with white-light illumination. **d**, Photoluminescence spectra obtained from the body of the nanowire (blue) and the end of the nanowire (green) at low pump power (10 mW). **e**, Spectrum from the nanowire end at higher pump power (80 mW) exhibiting periodic intensity variation. The period varies from 1.67 to 2.59 nm with increasing wavelength, which is consistent with the calculated mode spacing for the 18.8 μm nanowire and the dispersion of the refractive index, $n(\lambda)$ (ref. 25). Inset, end-emission intensity as a function of pump power. The nanowires in **d** and **e** were uniformly illuminated. **f**, Mode spacing versus inverse nanowire length. Green triangles, experimental points; red triangle, extrapolation to infinite length; red line, linear fit to these data. The contribution of $n(\lambda)$ was minimized by plotting the mode spacing at 530 nm in all cases. CdS nanowires were synthesized at 880 °C by laser-assisted catalytic growth^{12–14} using gold as the catalyst. The resulting nanowire product was dispersed in ethanol, and sonicated for 30–60 s to produce a high yield of wires with cleaved ends. Room- and low-temperature luminescence measurements were made with homebuilt epifluorescence microscopes, where a frequency-doubled Ti:sapphire laser (76 MHz, ~ 200 fs pulses, 410 nm) was used for optical excitation. Spectra were recorded using a 300 mm spectrometer (1200 lines mm^{-1} grating) and liquid-nitrogen-cooled charge-coupled device (CCD) detector. The room- and low-temperature instruments have spectral resolutions of 0.3 and 0.8 nm, respectively.

$(\lambda^2/2L)(n_1 - \lambda(dn_1/d\lambda))^{-1}$, where $dn_1/d\lambda$ is the dispersion relation for the refractive index. This expression provides a good description of the observed spacing when the measured nanowire length is equated with L , and moreover, analysis of similar data from nanowires of varying length demonstrates that the mode spacing is inversely proportional to the wire length (Fig. 1f), as expected. Together these results show that the CdS nanowires form a Fabry–Perot cavity. From the mode linewidths (background subtracted) we estimate a moderate cavity quality factor of about 600 at room temperature.

The observation of sharp modes in the uniform CdS nanowire gain medium in the superlinear regime is indicative of amplified spontaneous emission. Significantly, excitation at higher powers, which was possible in low-temperature experiments, leads to preferential gain in a single mode and the onset of lasing (Fig. 2a). Measurements of the linewidth dependence on pump power show an abrupt decrease to a value limited by our instrument resolution soon after the changeover to superlinear behaviour (Fig. 2b). In contrast, emission from the nanowire body is broad and approximately linearly dependent on excitation power, and the background spontaneous emission saturates in the superlinear regime (Supplementary Information). From the superlinear behaviour we estimate the threshold average pump power to be 40 kW cm^{-2} , although the threshold varies from nanowire to nanowire with the lowest value to date of around 2 kW cm^{-2} at low temperature.

These optical experiments demonstrate that individual nanowires can function as Fabry–Perot cavities and support lasing, although without electrical pumping nanowire lasers would be of limited technological importance^{1–3}. In general, electrically driven

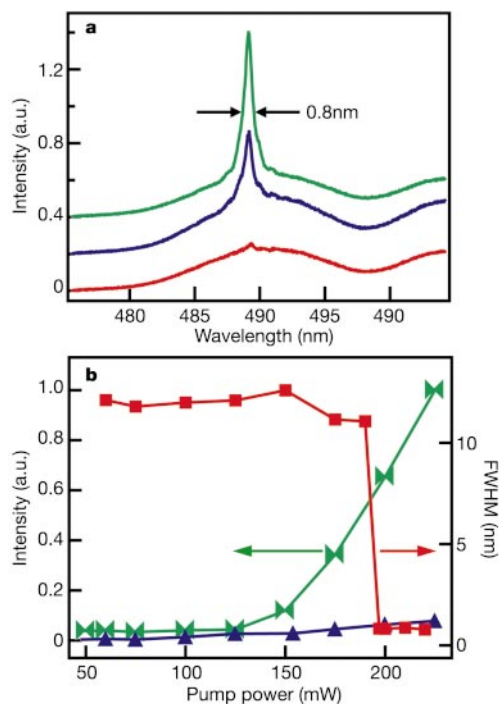


Figure 2 Optically pumped nanowire laser. **a**, Emission spectra from a CdS nanowire end with a pump power of 190, 197 and 200 mW (red, blue and green) recorded at 8 K. The spectra are offset by 0.2 intensity units for clarity. **b**, Emission intensity and FWHM of emission peaks versus laser pump power. The emission intensity from the nanowire body (blue) maintains a low value and is approximately linear in pump power, whereas the emission from the nanowire end (green) exhibits superlinear behaviour above 125 mW. The FWHM (red) has a nearly constant value of about 12 nm at low power, and abruptly narrows to the instrument-resolution value at high power. Solid symbols correspond to experimental data and lines are guides to the eye.

lasing requires efficient electron (n-type) and hole (p-type) injection into the cavity region. In the case of planar CdS structures, this has been difficult owing to problems in producing high-mobility p-type CdS or combining n-CdS with other high-mobility p-type materials. A clear advantage of nanowire-based structures is the ability to combine different high-quality materials as desired to achieve the required device function^{15–17}.

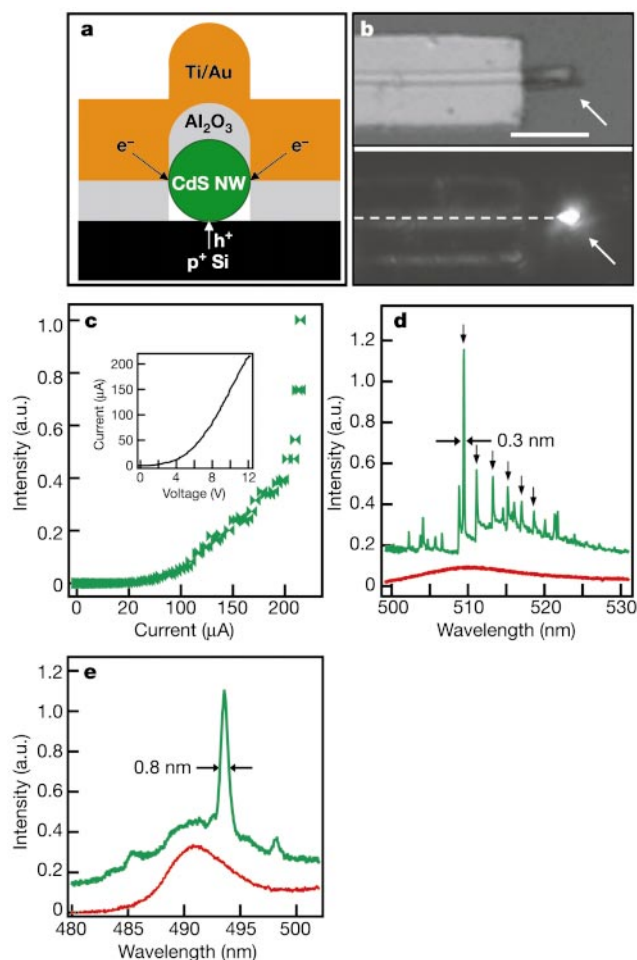


Figure 3 Nanowire electrical injection laser. **a**, Schematic showing the cross-section of the device structure. In this structure, electrons and holes can be injected into the CdS nanowire along the whole length from the top metal layer and the bottom p-Si layer, respectively. The devices were fabricated by assembling CdS nanowires on heavily doped p-Si on insulator substrates ($>4 \times 10^{19} \text{ cm}^{-3}$; 500 nm thick), followed by electron-beam lithography and electron-beam evaporation of 60–80 nm aluminium oxide, 40 nm Ti and 200 nm Au. One end of the nanowire was left uncovered for emission output from the device. **b**, Top panel shows an optical image of a device described in **a**. The arrow highlights the exposed CdS nanowire end. Scale bar, 5 μm. Bottom panel shows an electroluminescence image recorded from this device at room temperature with an injection current of about 80 μA. The arrow highlights emission from the CdS nanowire end. The dashed line highlights the nanowire position. **c**, Emission intensity versus injection current. The intensity increases rapidly above about 200 μA, which corresponds to the onset of lasing. Inset shows current versus voltage for this device. **d**, Electroluminescence spectra obtained from the nanowire end with injection currents of 120 μA (red) and 210 μA (green). The black arrows highlight Fabry–Perot cavity modes with an average spacing of 1.83 nm. The green spectrum is shifted upwards by 0.15 intensity units for clarity. **e**, Emission spectra from a CdS nanowire device with injection currents of 200 μA (red) and 280 μA (green) recorded at 8 K. The spectra are offset by 0.10 intensity units for clarity. The single peak observed at high injection has a linewidth of 0.8 nm, comparable to the instrument resolution and that observed in the optical pumping experiments.

Initial studies of electrical injection into CdS nanowire cavities were carried out using n-type CdS and p-type silicon (p-Si) crossed-nanowire structures^{15–17}. Transport studies of individual CdS nanowires show that they are n-type with doping concentrations on the order of 10^{18} – 10^{19} cm⁻³ and electron mobilities of about 100 cm² V⁻¹ s⁻¹ (Supplementary Information). Current–voltage (*I*–*V*) measurements made on a typical n-CdS/p-Si crossed nanowire structure show current rectification with a sharp forward-bias turn-on at about 2 V, consistent with the formation of a p–n diode (see Supplementary Information for band diagram of this heterostructure). In forward bias, these crossed-nanowire structures exhibit electroluminescence from the n-CdS/p-Si nanowire cross-point and stronger emission from the ends of the CdS nanowires, which is consistent with the CdS nanowires functioning as good waveguides. Electroluminescence spectra recorded from CdS nanowire ends exhibit a prominent modulation in the intensity that can be assigned to the longitudinal modes of nanowire Fabry–Perot cavities (our unpublished results), and thus these electroluminescence data are in agreement with our optically pumped data recorded (Fig. 1) from similar CdS nanowires.

To investigate nanowire injection lasers we have implemented a hybrid structure (Fig. 3a) in which the n-type CdS nanowire laser cavities are assembled onto p-Si electrodes defined in heavily p-doped planar substrates. This structure produces the n-CdS/p-Si heterojunction (Supplementary Information) needed for an injection device. The hybrid structure is analogous to the p–n diodes formed in the crossed n-CdS/p-Si nanowire devices, although in this case holes can be injected along the entire length of the CdS nanowire cavity in contrast to the single crosspoint in crossed-nanowire devices. An image of a typical device is shown in Fig. 3b. Current versus voltage data recorded from devices fabricated in this way show current rectification with a forward-bias turn-on of 2–5 V (inset, Fig. 3c and Supplementary Information), which is consistent with the formation of p–n diodes. The variation in turn-on voltage is believed to be due to Al₂O₃ barrier between the metal/CdS contact and/or oxide at the CdS/p-Si junction.

Images of the room-temperature electroluminescence produced in forward bias from these hybrid structures (Fig. 3b) exhibit strong emission from the exposed CdS nanowire ends. Measurements of the nanowire end electroluminescence intensity versus current (Fig. 3c) show an initial increase in the intensity at about 90 μ A and then a much more rapid and highly nonlinear increase at about 200 μ A. Significantly, at low injection currents the spectrum of the end emission (Fig. 3d) shows a broad peak with FWHM $\approx$ 18 nm, consistent with spontaneous emission, but above the 200 μ A threshold the spectrum quickly collapsed into a limited number of very sharp peaks with a dominant emission line at 509.6 nm (Fig. 3d). These sharp peaks emerge from only part of the spontaneous emission spectrum and had an average spacing of about 1.8 nm, which is consistent with the Fabry–Perot cavity modes for the length of the nanowire device. The observation of multiple peaks versus a single mode is often observed in studies of lasing in planar devices²⁰ pumped close to threshold. Other small, sharp peaks are also observed and their explanation will require more detailed consideration of the nanowire cavity. Lastly, the dominant mode has a linewidth limited by the instrument resolution of only 0.3 nm. Taken together, these observations provide strong evidence for lasing from single-nanowire injection structures at room temperature.

There are several additional points and experiments we have carried out that deserve comment. First, we believe that injection non-uniformity, which is due in part to non-ideal CdS/p-Si and metal/CdS junctions, limits these new lasers. For example, at present it is not possible to drive the devices substantially above the observed threshold for lasing, which would allow for better characterization of threshold behaviour and also would be expected to lead to single-mode output. Second, we have also carried out low-

temperature measurements on independent CdS injection laser devices. These data (Fig. 3e) clearly show that the spontaneous emission spectrum can collapse to a single instrument-resolution-limited mode that is characteristic of lasing; moreover, these results are very similar to the low-temperature optically pumped results (Fig. 2). Our studies suggest that with further improvements in nanowire laser assembly/fabrication, better optical quality and greater robustness will be possible. Third, the measurement geometry, in which emitted light is collected perpendicularly to the nanowire cavity axis, contributes to background signal above the lasing threshold²¹ and complicates determination of absolute current–intensity relationship for the nanowire lasers. We believe that future studies of these structures with emission collected in an end-on geometry could provide data of use for developing a quantitative understanding and improved performance of these injection lasers.

Our results show that nanoscale injection lasers can be made from single semiconductor nanowires, and describe a powerful approach for producing integrated electrically driven photonic devices. This basic approach, which relies upon bottom-up assembly of the key laser cavity/medium in a single step, can be extended to other materials, such as GaN²² and InP¹⁵ nanowires, to produce nanoscale lasers that not only cover the ultraviolet through near-infrared spectral regions but can also be integrated as single or multi-colour laser source arrays in silicon microelectronics and lab-on-a-chip devices. There are some scientific and technical challenges that may need to be addressed to realize this potential, such as the development of more efficient cavities and injection schemes. We believe that both issues could be addressed at the nanowire growth stage before device assembly by preparing Bragg gratings at the nanowire ends through axial composition modulation²³, and using core–shell nanowire structures²⁴ to enable uniform injection into the active medium/cavity, respectively. By addressing these and other issues, such as quantifying contributions to optical losses within the nanowire cavity, nanowire lasers could be developed into systems that affect several applications for solid-state lasers, including telecommunications and data storage, and may enable new applications in highly integrated chemical/biological sensors, near-field optical lithography, a host of scanning probe microscopies, and perhaps even laser-based surgery with unprecedented resolution. □

Received 1 November; accepted 29 November 2002; doi:10.1038/nature01353.

- Gray, G. R. in *Semiconductor Laser: Past, Present, and Future* (ed. Agrawal, G. P.) 284–320 (American Institute of Physics, New York, 1995).
- Kapon, E. (ed.) *Semiconductor Lasers I: Fundamentals* (Academic, San Diego, 1999).
- Kapon, E. (ed.) *Semiconductor Lasers II: Materials and Structures* (Academic, San Diego, 1999).
- Kozlov, V. G., Bulovic, V., Burrows, P. E. & Forrest, S. R. Laser action in organic semiconductor waveguide and double-heterostructure devices. *Nature* **389**, 362–363 (1997).
- Bergren, M., Dodabalapur, A., Slusher, R. E. & Bao, Z. Light amplification on organic thin films using cascade energy transfer. *Nature* **389**, 466–469 (1997).
- Tessler, N., Denton, G. J. & Friend, R. H. Lasing from conjugated-polymer microcavities. *Nature* **382**, 695–697 (1996).
- Hide, F. et al. Semiconductor polymers: A new class of solid state laser materials. *Science* **273**, 1833–1836 (1996).
- Klimov, V. I. et al. Optical gain and stimulated emission in nanocrystal quantum dots. *Science* **290**, 314–317 (2000).
- Kazes, M., Lewis, D. Y., Ebenstein, Y., Mokari, T. & Banin, U. Lasing from semiconductor quantum rods in a cylindrical microcavity. *Adv. Mater.* **14**, 317–321 (2002).
- Huang, H. M. et al. Room-temperature ultraviolet nanowire nanolasers. *Science* **292**, 1897–1899 (2001).
- Kozlov, V. G. et al. Structures for organic laser diode lasers and optical properties of organic semiconductor under intense optical and electrical excitations. *IEEE J. Quant. Elec.* **36**, 18–26 (2000).
- Morales, A. M. & Lieber, C. M. A laser ablation method for the synthesis of crystalline semiconductor nanowires. *Science* **279**, 208–211 (1998).
- Duan, X. & Lieber, C. M. General synthesis of compound semiconductor nanowires. *Adv. Mater.* **12**, 298–302 (2000).
- Gudiksen, M. S. & Lieber, C. M. Diameter selective synthesis of semiconductor nanowires. *J. Am. Chem. Soc.* **122**, 8801–8802 (2000).
- Duan, X., Huang, Y., Wang, J., Cui, Y. & Lieber, C. M. Indium phosphide nanowires as building blocks for nanoscale electronic and optoelectronic devices. *Nature* **409**, 66–69 (2001).
- Cui, Y. & Lieber, C. M. Functional nanoscale electronic devices assembled using silicon nanowire building blocks. *Science* **291**, 891–893 (2001).

17. Huang, Y. *et al.* Logic gates and computation from assembled nanowire building blocks. *Science* **294**, 1313–1317 (2001).
18. Chen, C.-L. *Elements of Optoelectronics and Fiber Optics* (Irwin, Chicago, 1996).
19. Wang, J., Gudiksen, M. S., Duan, X., Cui, Y. & Lieber, C. M. Highly polarized photoluminescence and polarization sensitive photodetectors from single indium phosphide nanowires. *Science* **293**, 1455–1457 (2001).
20. Bagnall, D. M., Ullrich, B., Sakai, H. & Segawa, Y. Micro-cavity lasing of optically excited CdS thin films at room temperature. *J. Cryst. Growth* **214/215**, 1015–1018 (2000).
21. Bagnall, D. M., Ullrich, B., Qiu, X. G., Segawa, Y. & Sakai, H. Microcavity lasing of optically excited cadmium sulphide thin films at room temperature. *Opt. Lett.* **24**, 1278–1280 (1999).
22. Huang, Y., Duan, X., Cui, Y. & Lieber, C. M. GaN nanowire nanodevices. *Nano Lett.* **2**, 101–104 (2002).
23. Gudiksen, G. S., Lauhon, L. J., Wang, J., Smith, D. & Lieber, C. M. Growth of nanowire superlattice structures for nanoscale photonics and electronics. *Nature* **415**, 617–620 (2002).
24. Lauhon, L. J., Gudiksen, M. S., Wang, D. & Lieber, C. M. Epitaxial core-shell and core-multishell nanowire heterostructures. *Nature* **420**, 57–61 (2002).
25. El Assali, K. *et al.* Some structural and optical properties of CdS thin films prepared by RF sputtering. *Phys. Status Solidi A* **178**, 701–708 (2000).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank H. Park, X. Zhuang & M. S. Gudiksen for helpful discussion. C.M.L. is grateful for support of this work by the Air Force Office of Scientific Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.M.L.
(e-mail: cml@cmliris.harvard.edu).

Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO₂

Robert M. DeConto* & David Pollard†

* Department of Geosciences, University of Massachusetts, Amherst, Massachusetts 01003, USA

† EMS Environment Institute, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

The sudden, widespread glaciation of Antarctica and the associated shift towards colder temperatures at the Eocene/Oligocene boundary (~34 million years ago) (refs 1–4) is one of the most fundamental reorganizations of global climate known in the geologic record. The glaciation of Antarctica has hitherto been thought to result from the tectonic opening of Southern Ocean gateways, which enabled the formation of the Antarctic Circumpolar Current and the subsequent thermal isolation of the Antarctic continent⁵. Here we simulate the glacial inception and early growth of the East Antarctic Ice Sheet using a general circulation model with coupled components for atmosphere, ocean, ice sheet and sediment, and which incorporates palaeogeography, greenhouse gas, changing orbital parameters, and varying ocean heat transport. In our model, declining Cenozoic CO₂ first leads to the formation of small, highly dynamic ice caps on high Antarctic plateaux. At a later time, a CO₂ threshold is crossed, initiating ice-sheet height/mass-balance feedbacks that cause the ice caps to expand rapidly with large orbital variations, eventually coalescing into a continental-scale East Antarctic Ice Sheet. According to our simulation the opening of Southern Ocean gateways plays a secondary role in this transition, relative to CO₂ concentration.

Antarctica has been located over southern polar latitudes since the

Early Cretaceous⁶, yet is thought to have remained mostly ice-free, vegetated, and with mean annual temperatures well above freezing until the Eocene/Oligocene boundary^{4,7}. Evidence for cooling and the sudden growth of an East Antarctic Ice Sheet (EAIS) comes from marine records (refs 1–3), in which the gradual cooling from the presumably ice-free warmth of the Early Tertiary to the cold 'icehouse' of the Late Cenozoic is punctuated by a sudden >1.0‰ rise in benthic δ¹⁸O values at ~34 million years (Myr). More direct evidence of cooling and glaciation near the Eocene/Oligocene boundary is provided by drilling on the East Antarctic margin⁸, the record of circum-Antarctic ice-rafted debris⁹, a shift in the clay composition of circum-Antarctic sediments¹⁰, and the fossil record of Antarctic vegetation¹¹. Glaciation is believed to have begun in the East Antarctic interior, discharging mainly via the Lambert Graben to Prydz Bay, with the Transantarctic Mountains restricting ice flow towards the Ross Sea until the ice sheets became larger in the Middle Oligocene⁴. Palaeogene Antarctic ice sheets appear to have been temperate, highly dynamic^{7,12}, and paced by Milankovitch orbital parameters^{1,13} in much the same way as the Quaternary ice sheets of the Northern Hemisphere.

The initial growth of the EAIS near the Eocene/Oligocene boundary has been attributed to the tectonic opening of ocean gateways between Antarctica and Australia (Tasmanian Passage), and Antarctica and South America (Drake Passage), leading to the organization of the Antarctic Circumpolar Current (ACC) and the 'thermal isolation' of Antarctica^{5,14}. This notion is supported by ocean general circulation model (OGCM) simulations, showing that the opening of Drake Passage and the organization of an ACC reduces southward oceanic heat transport and cools Southern Ocean sea surface temperatures (SSTs) by ~3 °C (refs 15, 16). However, although most tectonic reconstructions place the opening of the Tasmanian Passage close to the Eocene/Oligocene boundary, Drake Passage may not have provided a significant deep-water passage until several million years later^{17,18}. Additionally, these OGCM simulations lacked realistic atmospheric components, so the effects on Antarctic climate are unresolved.

Alternatively, declining atmospheric CO₂ may have played a

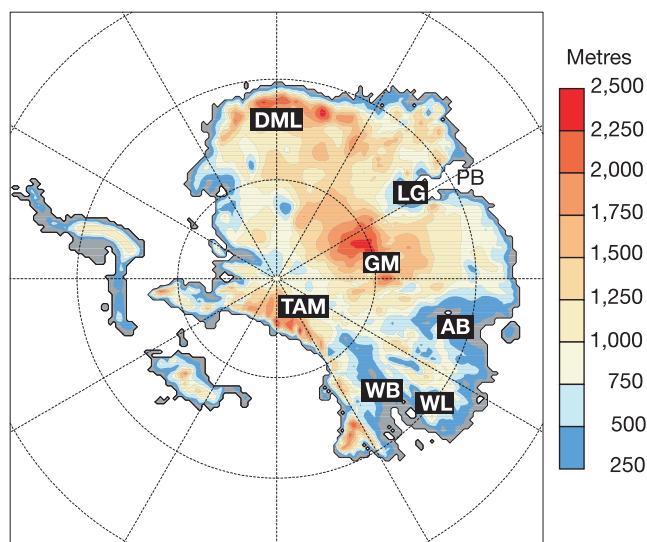


Figure 1 Early Cenozoic ice-free Antarctic topography in metres above sea level. This was reconstructed from a modern 5-km database³⁰, isostatically relaxed to ice-free equilibrium and interpolated to the 40-km polar stereographic grid of the ice-sheet model. Abbreviated place names mentioned in the text are also shown: AB, Aurora Basin; DML, Dronning Maud Land; GM, Gamburtsev Mountains; LG, Lambert Graben; PB, Prydz Bay; TAM, Transantarctic Mountains; WB, Wilkes Basin; WL, Wilkes Land.

fundamental role in Palaeogene cooling and Antarctic glaciation. Most estimates of early Cenozoic CO_2 mixing ratios range between 2 and 5 times present values, falling through the Cenozoic until reaching near-modern values sometime in the Neogene^{19,20}. The cooling due to declining p_{CO_2} would have gradually lowered annual snowline elevations until they intersected extensive regions of high Antarctic topography. Once some threshold was reached, feedbacks related to snow/ice-albedo and ice-sheet height/mass-balance^{21–23} could have initiated rapid ice-sheet growth during orbital periods favourable for the accumulation of glacial ice.

To test the relative importance of changing atmospheric CO_2 , orbital parameters and ocean heat transport in the nucleation and subsequent fluctuations of the early EAIS, we have developed a coupled global climate model (GCM) and dynamic ice-sheet model designed for experiments over long timescales. Previous dynamic ice-sheet models of ancient Antarctic ice sheets have used empirical parameterizations based on modern climates for their surface mass-balance forcing²⁴. Instead, we use the GENESIS (version 2.1) GCM²⁵

with a 50-m slab ocean (see Methods) responding to evolving Palaeogene boundary conditions to provide the meteorological input for our dynamic ice-sheet model experiments. A three-dimensional ice-sheet model of ice dynamics and bedrock response is asynchronously coupled to the GCM, allowing feedbacks between the ice sheet and atmospheric model components, and accounting for cyclical orbital forcing on 10^4 -yr timescales.

Boundary conditions for our Eocene/Oligocene climate/ice-sheet simulations are based on a global 34 Myr palaeogeography, including reconstructions of shorelines, rebounded ice-free topography (Fig. 1) and vegetation. To simulate the growth of Antarctic ice around the Eocene/Oligocene boundary, the GCM–ice-sheet model would ideally be run together through millions of years. However, the computational expense of the GCM makes that infeasible, so we have developed a new two-step scheme (see Methods) accounting for both orbital forcing and a decline in CO_2 , from $4 \times$ to $2 \times$ PAL (preindustrial atmospheric level, taken as 280 parts per million by volume (p.p.m.v.)) over a 10-Myr period.

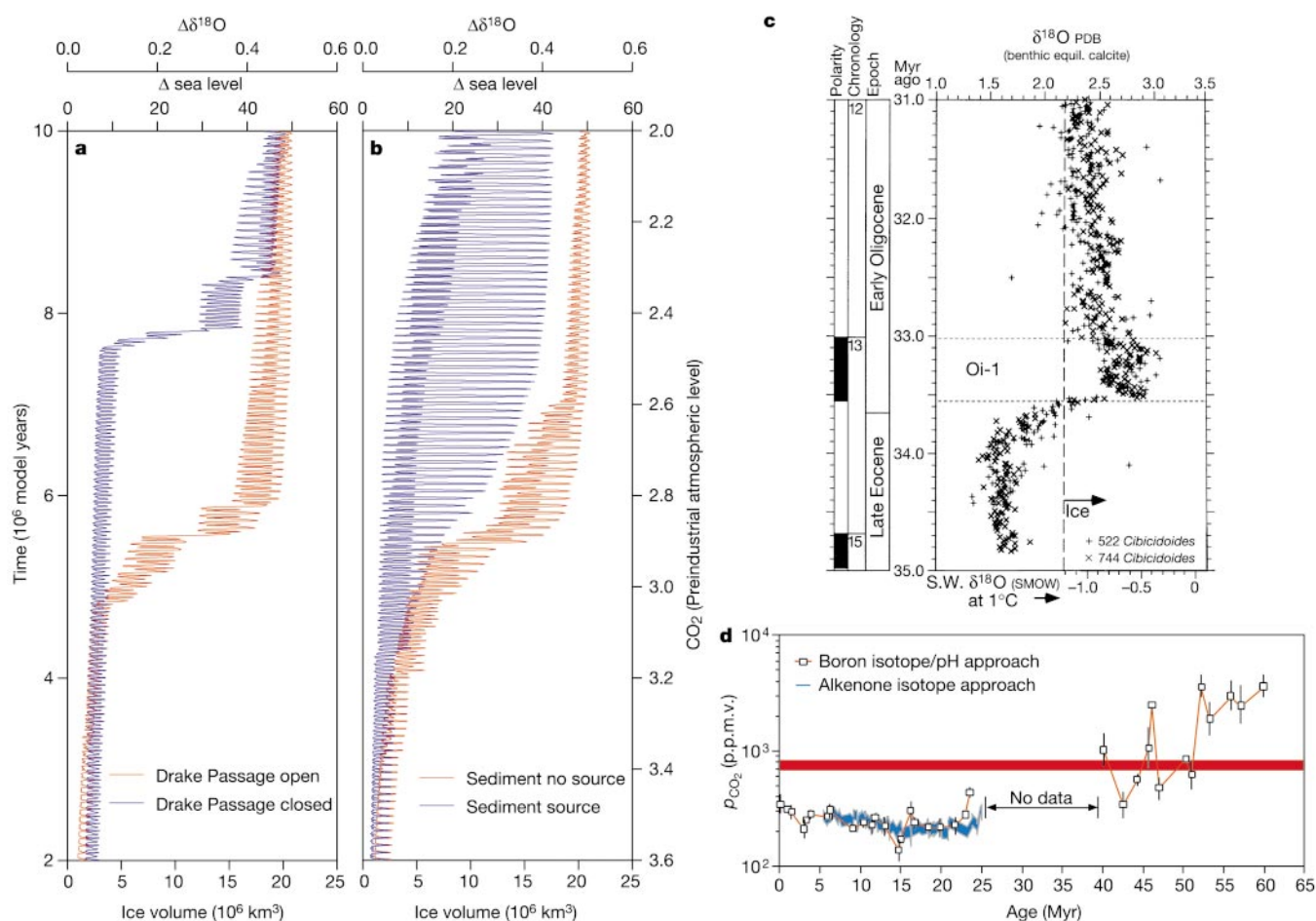


Figure 2 The transient climate-cryosphere response to a prescribed decline in CO_2 from $4 \times$ to $2 \times$ preindustrial atmospheric level over a 10-Myr period. Ice volume, and equivalent changes in sea level (Δ sea level) and the mean isotopic composition of the ocean ($\Delta\delta^{18}\text{O}$) are shown for two pairs of simulations: **a**, red curve, nominal case, representing an open Drake Passage; **a**, blue curve, with the global climate model ocean heat transport coefficient increased in the Southern Hemisphere to represent a closed Drake Passage; **b**, red curve, with deformable sediment added to the ice model and no generation of new till; **b**, blue curve, with the same sediment model but including subglacial generation of new till where basal ice is in contact with clean bedrock. Equivalent sea level changes were calculated according to the global ocean-area fraction in our 34-Myr palaeogeography (0.731). $\Delta\delta^{18}\text{O}$ was calculated assuming a 0.0091

change in $\delta^{18}\text{O}$ per 1-m change of sea level. **c**, High-resolution foraminiferal stable isotope data¹ from Ocean Drilling Program (ODP) sites 522 and 744 are shown for comparison with our simulated glacial transitions in **a** and **b**. The magnitude of the observed Eocene/Oligocene oxygen isotopic excursion is larger than modelled, probably because of the influence of deep-ocean cooling on $\delta^{18}\text{O}$, our assumed constant of proportionality with sea level being too low, and/or the model's lack of significant West Antarctic ice and its buttressing effect on the EAIS. **d**, proxy estimates of p_{CO_2} based on a boron isotope/pH approach¹⁹ and an alkenone isotope approach²⁰ reprinted from ref. 2 with the permission of the American Association for the Advancement of Science for comparison with the critical range of CO_2 values (thick red line) seen in **a** ($1 \times \text{CO}_2 = 280$ p.p.m.v.).

The 10-Myr integrations illustrate the possible evolution of the EAIS and several distinct modes of early ice-sheet variability as p_{CO_2} declined through the Palaeogene. In our nominal experiment (Fig. 2a, red), relatively small, isolated ice caps first form in the highest elevations of Dronning Maud Land in response to orographically lifted maritime moisture, producing high winter snowfall over the high elevations of the Polar Plateau/Gamburtsev Mountain region and along the Transantarctic Mountains (Fig. 3a). However, long-term ice variations are dominated by changes in summer temperatures and melting, not by changes in snowfall. At around $3 \times \text{CO}_2$, a small ice-sheet threshold is reached first for the Gamburtsev ice cap (Fig. 3b), which rapidly expands owing to height/mass-balance feedback^{21,22} during orbital periods with minimal summer insolation. The first significant ice caps are not large enough to reach sea level around much of the coast, but produce orbitally paced changes in ice volume large enough to account for significant variations in eustatic sea level (Fig. 2). Within the next several eccentricity cycles, additional jumps in ice volume occur as the individual ice caps permanently coalesce (Fig. 3c), remaining joined even through warm-summer orbital periods. At $\sim 2.8 \times \text{CO}_2$, the once individual ice caps form a continental-scale EAIS (Fig. 3d). As CO_2 continues to fall, ice-sheet extent becomes limited

more by continental shorelines and not by land-based ablation zones. This truncates the amplitude of the response to orbital forcing and increases ice-sheet stability for the remainder of the simulation (Fig. 2).

In our nominal 10 Myr experiment, the Drake and Tasman passages were effectively assumed to be open. To address the importance of the opening of Southern Ocean gateways on EAIS mass-balance, we adjusted our model to reflect changes found in previous OGCM sensitivity studies with an open versus closed Drake Passage^{15,16}. In these studies, a closed Drake Passage increases southward ocean heat transport by 15–20% over all southern latitudes. To mimic that effect while retaining the computational efficiency of a slab ocean, we adjusted the slab diffusion coefficient so as to increase ocean heat transport by 20% over the same latitudes, and repeated the 10-Myr experiment.

The two 10-Myr experiments in Fig. 2a have a permanently closed or open Drake Passage, allowing the effect on the timing of the CO_2 -induced transition to be seen clearly. The parameterized opening of Drake Passage has a significant effect on ice volume only when CO_2 values are within the critical range: $\sim 2.5 \times$ to $3 \times$ present. With Drake Passage closed, the transition to major glaciation is delayed until CO_2 falls an additional ~ 140 p.p.m.v. compared with that

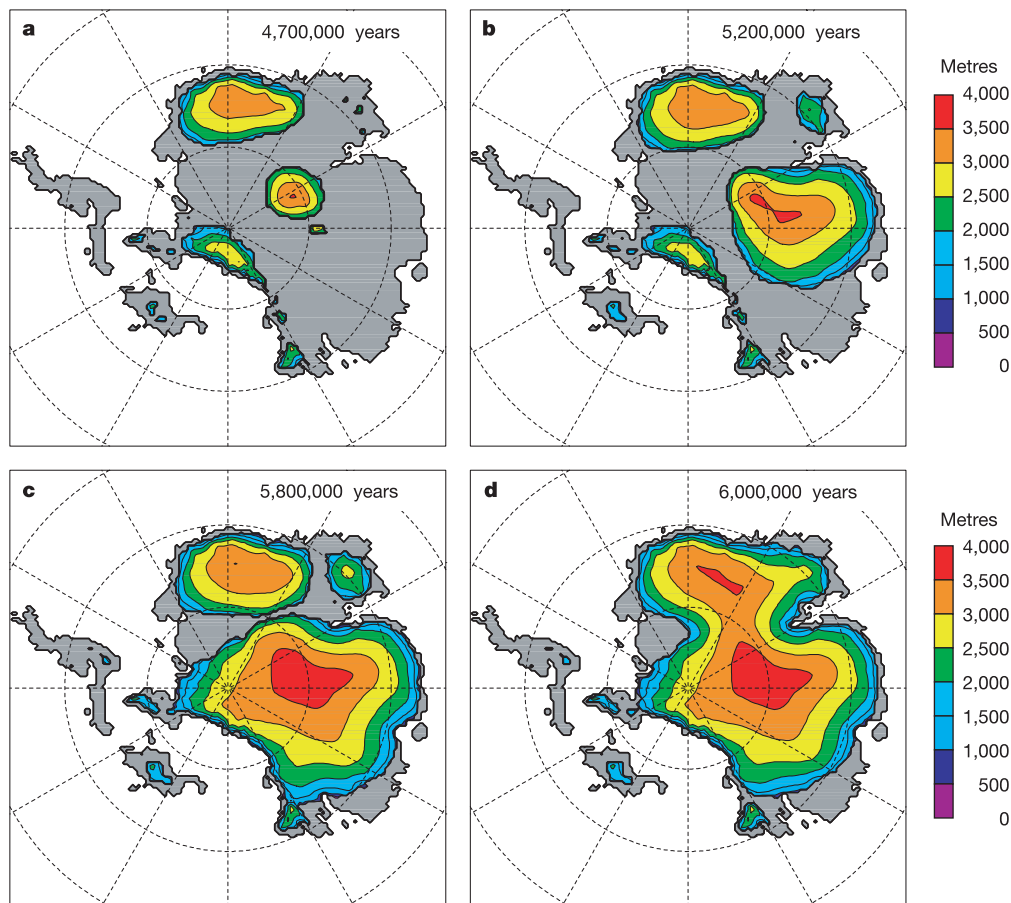


Figure 3 Ice-surface elevations at instantaneous times during the transition from ‘Greenhouse’ to ‘Icehouse’ conditions in our nominal 10-Myr simulation (Fig. 2a, red curve). **a**, At year 4,700,000, three small, isolated ice caps exist on the highest elevations of Dronning Maud Land and the Gamburtsev and Transantarctic mountains, fluctuating only slightly with orbital variations. **b**, At year 5,200,000 as CO_2 declines below $3 \times$ preindustrial atmospheric level, a small ice-sheet threshold^{21,22} has been reached for the Gamburtsev ice cap, which greatly expands in volume. **c**, At year 5,800,000, further decline of CO_2 allows the Gamburtsev and Transantarctic ice caps to coalesce

permanently, buttressing each other to form a much larger central ice sheet. This central ice cap makes repeated, but temporary, contact with the Dronning Maud Land ice cap during orbital periods with minimal summer insolation. **d**, By year 6,000,000, a further jump in volume occurs as the Dronning Maud ice cap permanently coalesces with the rest of the ice mass, forming a continental-scale East Antarctic Ice Sheet with two main spreading centres and major discharges down the Lambert Graben and other bedrock troughs to Prydz Bay and the Wilkes and Aurora basins.

when Drake Passage is open. Thus, a 20% change in southward oceanic-heat convergence has a significant effect on the timing of glacial inception, but only within a relatively narrow range of CO_2 (Fig. 2). If Drake Passage opened outside this range of heightened cryosphere sensitivity, the effect would be small. If the opening occurred at levels of CO_2 greater than $\sim 3 \times$ present, a continental ice sheet would not have formed. If the opening of Drake Passage occurred after 34 Myr ago, as suggested by geologic evidence¹⁸, then the EAIS ice sheet already in existence would have become only slightly larger.

Another 10-Myr simulation was done with the addition of a subglacial deforming sediment component in the ice-sheet model. Subglacial processes are thought to be important for the behaviour of fast-moving West Antarctic ice streams²⁶ and the evolution of Northern Hemisphere Quaternary ice sheets²⁷. The sediment model²⁷ accounts for extra sliding within a sediment layer of variable thickness, and predicts the evolving continental-scale distribution of till (see Methods). The addition of sediment in the 10-Myr integration (Fig. 2b, red) results in enhanced orbital variability, especially around the major transition, but it does not significantly affect the transition's timing. The enhanced variability is caused by thinner ice-sheet profiles and wider ablation zones, as found for Quaternary ice sheets of the Northern Hemisphere²⁷. After the transition, the ice sheets eventually revert to the behaviour of the non-sediment model, after easily eroded sediment in the continental interior is scraped clean by repeated cycles. When the sediment model is modified to include subglacial generation of new till (see Methods), considerable amounts of sediment are maintained under the ice and the orbital variability remains much larger through most of the simulation (Fig. 2b, blue). Compared with proxy records influenced by ice volume across the Eocene/Oligocene boundary (Fig. 2c), the simulations with no till generation (Fig. 2b, red) or no sediment at all (Fig. 2a, red) appear more realistic. However, this simple representation of continental-scale sediment is intended only to illustrate the possible envelope of effects on Antarctic ice evolution.

In summary, our model results provide an alternative explanation for the sudden build-up and subsequent variations of Antarctic ice near the Eocene/Oligocene boundary, emphasizing atmospheric CO_2 , orbital forcing, and ice-climate feedbacks as the primary causes of the Eocene/Oligocene climate transition. As CO_2 declined through the Palaeogene, the gradual lowering of the Antarctic snowline began intersecting areas of high topography, first producing small, isolated ice caps. As CO_2 continued to fall, height/mass-balance feedbacks were initiated suddenly, producing larger dynamic ice sheets that alternately coalesced and separated in response to orbital forcing. With further decline of CO_2 , a single, large EAIS became a more permanent feature, almost insensitive to orbital forcing, with very little summer melting, and accumulation zones reaching sea-level around most of the continent. Although the rate of CO_2 decline used in our model is only 56 p.p.m.v. per million years, it does produce a rather sudden climatic transition due to the initiation of internal climate system feedbacks. These results are in good agreement with high-resolution benthic $\delta^{18}\text{O}$ and Mg/Ca data spanning the Eocene/Oligocene boundary (refs 1 and 3) (Fig. 2c) and the sudden occurrence of ice-rafted debris in the Southern Ocean⁹. Although the simulated CO_2 /ice-sheet initiation threshold (between $3 \times$ and $2.5 \times$ PAL) may be model dependent, it is within the general range of proxy estimates of mid-Cenozoic p_{CO_2} (refs 19 and 20) (Fig. 2d).

The opening of Southern Ocean gateways and the formation of the ACC undoubtedly cooled high southern latitudes. However, the simulated effect of a 20% change in ocean heat transport associated with the opening of Drake Passage is shown to have a smaller effect than that expected in the transition from a 'greenhouse' to an 'icehouse' climate. In our model, the opening of Drake Passage can only be a potential trigger for glacial inception when atmospheric

CO_2 is within a relatively narrow range (Fig. 2a, d), reinforcing the importance of p_{CO_2} as a fundamental boundary condition for Cenozoic climate change. More explicit modelling of changes in ocean circulation using coupled atmospheric-oceanic GCMs, along with more proxy estimates of Palaeogene CO_2 around the Eocene/Oligocene boundary, will help to reveal the relative role of climate-forcing factors on Cenozoic Antarctic glaciation. □

Methods

Global climate model

The atmospheric component of GENESIS version 2.1 has 18 vertical layers and a spectral resolution of T31 ($\sim 3.75^\circ \times 3.75^\circ$), coupled to $2^\circ \times 2^\circ$ surface models including a non-dynamic 50-m slab ocean, a dynamic-thermodynamic sea-ice model, and multi-layer models of soil, snow and vegetation. Ocean heat transport is not prescribed, but is calculated as linear diffusion down the local ocean-temperature gradient, with the coefficient depending on latitude and the zonal fraction of land versus sea. Modifications to the standard GENESIS version 2.0 (ref. 25) include an improved slab diffusion coefficient, which yields realistic present zonal ocean heat transport.

Ice-sheet model

The thermo-mechanical ice-sheet model is based on the vertically integrated continuity equation for ice mass (refs 24 and 29), using a $40 \text{ km} \times 40 \text{ km}$ polar stereographic grid. The evolution of ice geometry is determined by surface mass balance, basal melting, and ice flow. Ice temperatures are predicted to account for their effect on rheology and basal sliding. The local bedrock response to ice load is a simple relaxation towards isostasy, with a time constant of 5,000 years. Lithospheric flexure is modelled by linear elastic deformation²⁸. Ice shelves are not included, and ice extent is constrained by continental shorelines. This ignores sea-level change and inhibits the formation of a marine-based West Antarctic Ice Sheet, which will be addressed in future work.

Deforming sediment model

The sediment component added to the ice-sheet model accounts for extra sliding due to horizontal shear in a layer of subglacial sediment, driven by the basal shear stress of the ice. The most appropriate sediment rheology is controversial²⁶; we use a slightly nonlinear rheology²⁷, extended here to two horizontal dimensions. This allows significant shear in the upper tens of centimetres to approximately 1 m of sediment below the ice, with deformation occurring only where the basal temperature is at the pressure melting point and the sediment is saturated. Our 10-Myr simulations are initialized with a uniform 50 m layer of regolith, assumed to have accumulated before the onset of major glaciation. Two simple parameterizations are added to simulate the long-term continental-scale evolution of sediment: (1) ice-free transport of moraines by rivers, modelled as instantaneous fluvial transport down the topographic gradient to the coast, lakes, or ice dams; and (2) subglacial generation of new till, proportional to the work done by basal sliding where the existing sediment is thin or non-existent.

Global climate model/ice-sheet coupling

Relevant monthly mean meteorological fields used in the calculation of net annual surface-ice mass balance (surface air temperature and precipitation) are horizontally interpolated from the GCM to the much finer ice-sheet grid. Temperatures are lapse-rate adjusted to account for the GCM's coarse representation of topography²⁵. A computationally economical positive degree-day parameterization with an imposed diurnal cycle is used to calculate accumulation and ablation from the monthly mean temperatures and precipitation²⁹, including a correction for refreezing of meltwater²⁵. Mass balance is re-computed every 200 years to account for changing ice elevations.

First step of long-term integrations

To perform long (10^6 yr) ice-sheet simulations and still make use of GCM climate solutions, we use a two-step asynchronous coupling scheme. In the first step, relatively short, 40 kyr simulations are performed with varying orbital parameters but fixed CO_2 . A synthetic orbital sequence is used (see Supplementary Information), with periodicities that are exact multiples or divisors of 40 kyr; that is, precession, obliquity and eccentricity vary sinusoidally with periods of 20 kyr, 40 kyr and 80 kyr, respectively. This captures the essence of orbital forcing. In each 40-kyr run, the ice-sheet model is integrated continuously, starting with no ice. The GCM is run for a few decades at the start of each simulation, and again every 10 kyr with updated ice-sheet size and orbit parameters to provide continuous mass balance for the ice model. Monthly mean temperatures and precipitation from the last 10 yr of each of the five GCM snapshots are stored for later use. Four 40-kyr sequences were conducted, with fixed CO_2 levels of $2 \times$ and $3 \times$ PAL, with and without a GCM modification representing Drake Passage closure (see text).

Second step of long-term integrations

The ice-sheet model is driven continuously for 10 million years, computing mass balances every 200 years from the appropriate linear weights of the stored sets of monthly GCM climates from the first step. The same synthetic 40-kyr orbital sequence is repeated in alternating forward and backward passes, to produce a smooth 80-kyr eccentricity cycle. The ice-driving climatology is linearly weighted in time between the two surrounding orbital GCM solutions. In addition, a steady decline in atmospheric CO_2 is specified, from $4 \times$ to $2 \times$ PAL over 10 Myr, by interpolating/extrapolating with respect to CO_2 level

between the appropriate GCM climates and accounting for the logarithmic radiative effect of CO₂. Unlike the shorter 40-kyr simulations (first step), ice albedo and topographic feedbacks are not calculated explicitly, but the latter are captured by the lapse-rate correction in the mass-balance calculations.

Received 25 July; accepted 12 November 2002; doi:10.1038/nature01290.

1. Zachos, J. C., Quinn, T. M. & Salamy, K. A. High-resolution (10⁴ years) deep-sea foraminiferal stable isotope records of the Eocene–Oligocene climate transition. *Paleoceanography* **11**, 251–266 (1996).
2. Zachos, J., Pagani, M., Sloan, L. & Thomas, E. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
3. Lear, C. H., Elderfield, H. & Wilson, P. A. Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite. *Science* **287**, 269–272 (2000).
4. Barrett, P. J. Antarctic paleoenvironment through Cenozoic times—a review. *Terr. Antarct.* **3**, 103–119 (1996).
5. Kennett, J. P. Cenozoic evolution of Antarctic glaciation, the circum-Antarctic oceans and their impact on global paleoceanography. *J. Geophys. Res.* **82**, 3843–3859 (1977).
6. Lawver, L. A., Gahagan, L. M. & Coffin, M. F. in *The Antarctic Paleoenvironment: A Perspective on Global Change* (eds Kennett, J. P. & Warnke, D. A.) 7–30 (American Geophysical Union, Washington DC, 1992).
7. Hambrey, M. J., Larsen, B. & Ehrmann, W. U. in *Ocean Drilling Program Scientific Results 119* (eds Barron, J. & Larsen, B.) 77–132 (College Station, Texas, 1991).
8. Wilson, G. S., Roberts, A. P., Verosub, K. L., Florindo, F. & Sagnotti, L. Magnetobiostratigraphic chronology of the Eocene–Oligocene transition in the CIROS-1 core, Victoria Land margin, Antarctica: Implications for Antarctic glacial history. *Geol. Soc. Am. Bull.* **110**, 35–47 (1998).
9. Zachos, J. C., Breza, J. R. & Wise, S. W. Early Oligocene ice sheet expansion on Antarctica: stable isotope and sedimentological evidence from Kerguelen Plateau, southern Indian Ocean. *Geology* **20**, 569–573 (1992).
10. Ehrmann, W. U. & Mackensen, A. Sedimentologic evidence for the formation of an East Antarctic ice sheet in Eocene/Oligocene time. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **93**, 85–112 (1992).
11. Francis, J. E. Evidence from fossil plants for Antarctic Paleoclimates over the past 100 million years. *Terr. Antarct. Rep.* **3**, 43–52 (1999).
12. Barrett, P. J., Elston, D. P., Harwood, D. M., McKelvey, B. C. & Webb, P.-N. Mid-Cenozoic record of glaciation and sea-level change on the margin of Victoria Land basin, Antarctica. *Geology* **15**, 634–637 (1987).
13. Naish, T. R. *et al.* Orbitally induced oscillations in the East Antarctic ice sheet at the Oligocene/Miocene boundary. *Nature* **413**, 719–723 (2001).
14. Exon, N., Kennett, J., Malone, M. & the Leg 189 Shipboard Scientific Party. The opening of the Tasmanian gateway drove global Cenozoic paleoclimatic and paleoceanographic changes: results of Leg 189. *JOIDES J.* **26**, 11–17 (2000).
15. Toggweiler, J. R. & Bjornsson, H. Drake Passage and paleoclimate. *J. Quat. Sci.* **15**, 319–328 (2000).
16. Nong, G. T., Najjar, R. G., Seidov, D. & Peterson, W. Simulation of ocean temperature change due to the opening of Drake Passage. *Geophys. Res. Lett.* **27**, 2689–2692 (2000).
17. Lawver, L. A. & Gahagan, L. M. in *Tectonic Boundary Conditions for Climate Reconstructions* (eds Crowley, T. J. & Burke, K. C.) 212–223 (Oxford Univ. Press, New York, 1998).
18. Barker, P. F. & Burrell, J. The opening of Drake Passage. *Mar. Geol.* **25**, 15–34 (1977).
19. Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide over the past 60 million years. *Nature* **406**, 695–699 (2000).
20. Pagani, M., Arthur, M. A. & Freeman, K. H. Miocene evolution of atmospheric carbon dioxide. *Paleoceanography* **14**, 273–292 (1999).
21. Birchfield, G. E., Weertman, J. & Lunde, A. T. A model study of the role of high latitude topography in the climatic response to orbital insolation anomalies. *J. Atmos. Sci.* **39**, 71–87 (1982).
22. Abe-Ouchi, A. & Blatter, H. On the initiation of ice sheets. *Ann. Glaciol.* **18**, 203–207 (1993).
23. Maqueda, M., Willmott, A. J., Bamber, J. L. & Darby, M. S. An investigation of the small ice cap instability in the Southern Hemisphere with a coupled atmosphere–sea ice–ocean–terrestrial ice model. *Clim. Dyn.* **14**, 329–352 (1998).
24. Huybrechts, P. Glaciological modelling of the late Cenozoic East Antarctic ice sheet: stability or dynamism? *Geograf. Annal.* **75**, 221–238 (1993).
25. Thompson, S. L. & Pollard, D. Greenland and Antarctic mass balances for present and doubled atmospheric CO₂ from the GENESIS Version-2 Global Climate Model. *J. Clim.* **10**, 871–900 (1997).
26. Kamb, B. in *The West Antarctic Ice Sheet: Behaviour and Environment* (eds Alley, R. A. & Bindshadler, R. A.) 157–199 (American Geophysical Union, Washington DC, 2001).
27. Clark, P. U. & Pollard, D. Origin of the mid-Pleistocene transition by ice-sheet erosion of regolith. *Paleoceanography* **13**, 1–9 (1998).
28. Brothie, J. F. & Sylvester, R. On crustal flexure. *J. Geophys. Res.* **74**, 5240–5252 (1969).
29. Ritz, C., Fabre, A. & Letreguilly, A. Sensitivity of a Greenland ice-sheet model to ice flow and ablation parameters: consequences for the evolution through the last climate cycle. *Clim. Dyn.* **13**, 11–24 (1997).
30. Bamber, J. A. & Bindshadler, R. A. An improved elevation dataset for climate and ice-sheet modelling: validation with satellite imagery. *Ann. Glaciol.* **25**, 439–444 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements This material is based upon work supported by the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.D. (e-mail: deconto@geo.umass.edu).

Evolution of the Archaean crust by delamination and shallow subduction

Stephen F. Foley*, Stephan Buhre† & Dorrit E. Jacob*

* Institut für Geologische Wissenschaften, Universität Greifswald, F.L.-Jahnstrasse 17a, 17487 Greifswald, Germany

† Institut für Mineralogie, Universität Frankfurt, Senckenberganlage 28, 60054 Frankfurt-am-Main, Germany

The Archaean oceanic crust was probably thicker than present-day oceanic crust owing to higher heat flow and thus higher degrees of melting at mid-ocean ridges¹. These conditions would also have led to a different bulk composition of oceanic crust in the early Archaean, that would probably have consisted of magnesium-rich picrite (with variably differentiated portions made up of basalt, gabbro, ultramafic cumulates and picrite). It is unclear whether these differences would have influenced crustal subduction and recycling processes, as experiments that have investigated the metamorphic reactions that take place during subduction have to date considered only modern mid-ocean-ridge basalts^{2,3}. Here we present data from high-pressure experiments that show that metamorphism of ultramafic cumulates and picrites produces pyroxenites, which we infer would have delaminated and melted to produce basaltic rocks, rather than continental crust as has previously been thought. Instead, the formation of continental crust requires subduction and melting of garnet-amphibolite⁴—formed only in the upper regions of oceanic crust—which is thought to have first occurred on a large

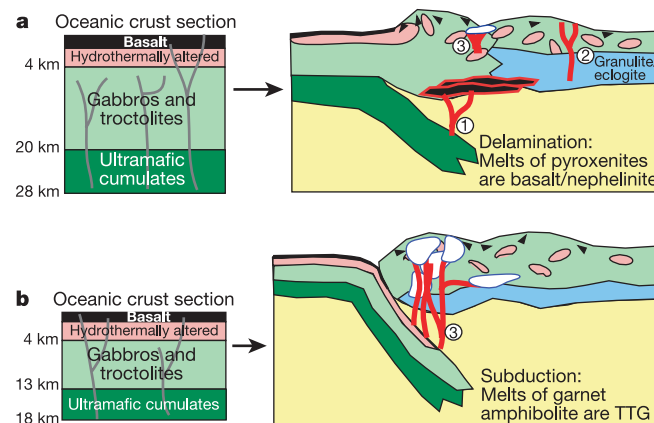


Figure 1 Model for the constitution and fate of oceanic crust. **a**, Hadean/early Archaean; delamination of ultramafic cumulates occurs. **b**, Late Archaean; subduction of the whole ocean crust occurs. Oceanic crust sections are estimated on the basis of eclogite xenolith petrology and chemistry^{12–14}, and thermodynamic modelling of the fractionation of picrites¹⁵. Parental picrites fractionate at the crust–mantle boundary to produce about 30% ultramafic cumulates, predominantly olivine and clinopyroxene. Grey lines indicate that more picritic melts are thought to have approached the surface in the Archaean^{13,16}. Melts may be produced by three main processes: in the Hadean and early Archaean, the metamorphosed ultramafic cumulates (now pyroxenites—see Fig. 2c) delaminate and melt to produce basaltic to nephelinitic melts which are greatly different from early continental crust (process 1). The lower crust may melt as eclogite¹⁰ (process 2), but only after removal of the lower ultramafic cumulate layer. The production of melts from garnet-amphibolites (process 3) is needed to produce continental-crust-like TTG suite⁴; this must have been rare (although not absent) in the early Archaean before the subduction of the hydrothermally altered uppermost basaltic crustal layer, but became important once the whole crust could be subducted²⁰.

scale during subduction in the late Archaean⁵. We deduce from this that shallow subduction and recycling of oceanic crust took place in the early Archaean, and that this would have resulted in strong depletion of only a thin layer of the uppermost mantle. The misfit between geochemical depletion models and geophysical models for mantle convection (which include deep subduction) might therefore be explained by continuous deepening of this depleted layer through geological time^{5,6}.

No unequivocal samples of Archaean oceanic crust are preserved at Earth's surface, so that processes surrounding the subduction and recycling of Archaean oceanic crust remain obscure. Indirect clues come from the preserved early continental crust, which broadly corresponds to tonalite–trondhjemite–granodiorite gneisses (TTG)⁷. These gneisses formed by partial melting of either eclogitic or amphibolitic oceanic crust^{8–10}. However, recent trace-element partitioning results show that the paired low Nb/Ta ratio and high Zr/Sm ratio of TTG must have formed by melting of garnet–amphibolite, and not eclogite⁴. The occurrence and melting of eclogite and/or garnet–amphibolite in the oceanic crust is therefore of central importance to assessing Archaean processes. Here, we consider where and when melting of garnet–amphibolite could have occurred in the Archaean oceanic setting.

The interpretation of greenstone belts as oceanic crust is not uncommon, but controversial¹, and if oceanic, they may be formed by the accretion of marginal basins and not represent crust of the open ocean¹¹. We prefer to consider eclogite xenoliths carried to the surface by kimberlites, as these may provide the only direct samples of crust formed in Archaean ocean basins. Many of these eclogites may be of late Archaean age, judging by the age determinations currently available (3.0–2.5 Gyr; refs 12–14). Those representing former volcanic rocks are picrites with more MgO than modern mid-ocean-ridge basalts (MORB; 5–8%) but considerably less MgO than komatiites. We interpret this to mean that the bulk oceanic crust was not komatiitic at any time after the Hadean era.

Figure 1 presents a model for the Archaean oceanic crust based on evidence from the eclogite suites and on thermodynamic calcu-

lations of the fractionation of picritic melts¹⁵. Owing to continual cooling of the Earth, the oceanic crust thinned from in excess of 25 km at about 3.5 Gyr ago to less than 20 km at 2.5 Gyr ago. The best modern analogues for such thick crust are oceanic plateaux^{10,15}, but these are not directly comparable as they form by near-stationary plume activity and not at spreading ridges. Oceanic plateaux are modelled as comprising ultramafic cumulates and basalt–gabbro–troctolite fractionates in a thickness ratio of about 1:3 from picritic parental melts¹⁵. Thus, an 18-km-thick crust would have 12–13 km basalt and gabbro above ultramafic cumulates consisting mostly of olivine and pyroxene¹⁵ (Fig. 1). However, recent studies of high-level gabbros in the oceanic crust show that primitive melts can approach the surface in modern fast-spreading ridges¹⁶, so that more rapid extension in the Archaean may have led to more picritic components in the upper parts of the crust than on oceanic plateaux today. Many eclogite xenoliths are the metamorphic products of gabbro protoliths, in agreement with fractionation models showing that basaltic residual melts should be widespread¹⁵.

Furthermore, ridge melts generally have very low water contents, as do plume-related melts¹⁷, so that the formation of amphibolite on metamorphism of basaltic fractionates would be limited to the uppermost hydrothermally altered oceanic crust, probably 3–4 km depth¹. Below this, granulites and eclogites would dominate. An important time constraint is that TTG melt production requires melting of the hydrothermally altered uppermost oceanic crust as garnet–amphibolite⁴.

The behaviour of modern MORB during subduction metamorphism is well studied experimentally^{2,3}, whereas no studies on picrites have been undertaken (to our knowledge). Therefore, we have conducted high-pressure experiments on a lava from Gorgona island, Colombia, with 17.5 wt% MgO (Table 1). MORB transforms with increasing pressure from granulite to garnet–granulite at 1.0–1.2 GPa depending on temperature, and then to eclogite at 1.5–1.7 GPa owing to the elimination of plagioclase. If wet, these rocks correspond to amphibolite, garnet–amphibolite and eclogite

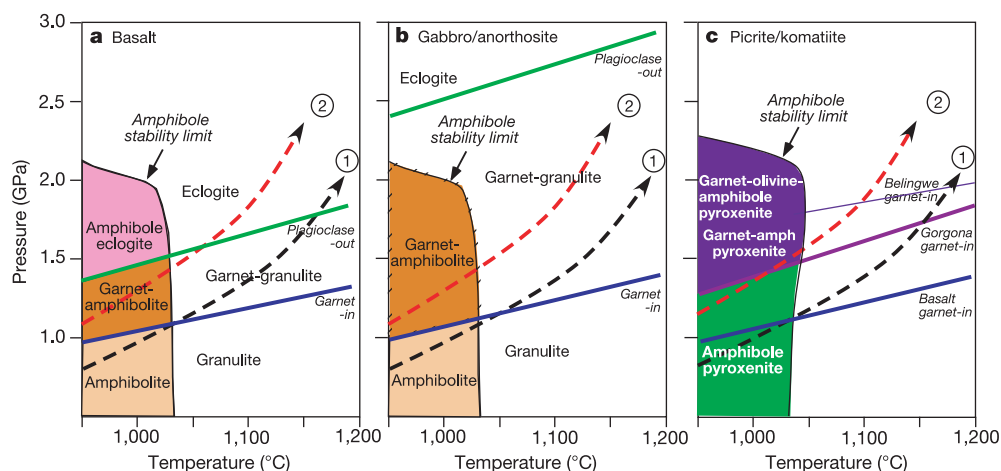


Figure 2 Stability of metamorphic products of parts of the Archaean oceanic crust. Coloured regions represent rocks produced by subduction metamorphism of **a**, MORB^{2,3}, **b**, gabbro/anorthosite, corresponding to plagioclase-rich plutonic rocks², and **c**, the Gorgona composition studied here. Garnet is stable above the 'garnet-in' curves, which lie at similar pressures for MORB and anorthosite, but at higher pressures with increasing MgO content for Gorgona and Belingwe komatiite (26.5% MgO; S.B., unpublished data). Garnet–granulite has widely differing stability fields in **a** and **b**, but does not occur at all in **c**. Melting points do not change much as a function of the MgO content of the rock wherever amphibole is present (phase relations interpolated after refs 29 and 30); melting occurs where the geotherms (dashed lines showing schematically conditions for early (geotherm 1) and late (geotherm 2) Archaean) pass out of the (coloured) regions of

subsolidus rocks. The Gorgona rock (**c**) melts along subduction geotherm 1 as amphibole–pyroxenite (dark green)—neither garnet nor plagioclase are stable—and along geotherm 2 as a garnet and amphibole-bearing pyroxenite (purple). The metamorphic equivalents of picrites and ultramafic cumulates are too MgO-rich to contain plagioclase during subduction. Formation of TTG gneisses requires melting of garnet–amphibolite⁴, which occurs for both basalt and gabbro/anorthosite (**a**, **b**), but only after cooling to conditions colder than geotherm 1. This prevents widespread formation of TTG with low Nb/Ta in the early Archaean; this must await a general thinning of the oceanic crust until the hydrothermally altered basaltic upper levels can be subducted and transform to amphibolite. No TTG-like melts can be produced by melting picrites or ultramafic cumulates under any conditions (**c**).

at lower temperatures (Fig. 2a). For plagioclase-rich cumulates, the fields for garnet- and plagioclase-bearing rocks extend to much higher pressures, but begin at about the same depth (Fig. 2b). Our results show that the metamorphic products of picrites or ultramafic cumulates (dark green and grey in Fig. 1) are plagioclase-free, and that the incoming of garnet lies at higher pressures than for MORB (1.7 GPa at 1,100 °C; Fig. 2c). These rocks can never transform to eclogite or garnet-amphibolite under subsolidus conditions, but will transform to rocks bearing orthopyroxene and/or olivine. Water-bearing equivalents are amphibole–pyroxenite, garnet-amphibole–pyroxenite, and garnet-olivine-amphibole–pyroxenite with increasing pressure (Fig. 2c). Partial melts of none of these pyroxenites will correspond to TTG or any other continental-crust-like composition, but will produce melts that are essentially basaltic, even if wet¹⁸. Also, these rocks may not undergo the sudden increase in density experienced by basalts during the granulite–eclogite transformation².

Subduction geotherms during the Archaean era are very poorly known, but will have been disproportionately hotter than modern ones owing to the dominance of smaller plates that were younger and hotter on subduction than today¹⁹. The eclogite xenoliths have very low Nb/Ta ratios indicating strong source depletion, and are reminiscent of a 'subduction zone' geochemical signature¹³ fitting a model with small plates and common ridge–subduction interaction¹⁹. Schematic geotherms are shown in Fig. 2; geotherm 1 would apply to the early Archaean, whereas geotherm 2 represents conditions in the latest Archaean. The following tectonic model is consistent with our experiments and the melting of other rock types in subduction zones, and with the production rates of TTG and early continental crust⁵. In the early Archaean, the crust was too thick to be subducted as a unit²⁰, and so the lower parts were delaminated and melted (Figs 1 and 2c). Melts of these metamorphosed ultramafic cumulates, now pyroxenites (Fig. 2c), are basaltic to nephelinitic¹⁸, and so the production of voluminous TTG-like melts was not favoured in Archaean subduction-like settings.

The oceanic crust became steadily thinner until, in the late Archaean (3.3–3.0 Gyr ago), a threshold was reached after which the whole crust could be subducted²⁰. At this time, the hydrothermally altered uppermost 4 km of crust was introduced into the subduction zone, and basaltic rocks were subject to widespread melting as garnet-amphibolite (Fig. 2a, b). Now, large volumes of TTG-like melts with high Zr/Sm and low Nb/Ta (ref. 4) could be produced for the first time, explaining the increase in volume of continental crust preserved from the late Archaean⁵. TTG melts could be produced during the delamination period of the early Archaean (Fig. 1) only from eclogitic lower crust from which the ultramafic cumulates had already been removed (process 2 in Fig. 1),

or by melting of amphibolite within the thick crust (process 3 in Fig. 1). Partial melting of eclogite is commonly invoked to produce TTG¹⁰, but would have to await delamination of the cumulate pile and so removal of the thermal protection from mantle temperatures. Furthermore, melting of eclogites in the lower parts of thick oceanic crust would produce rocks of broadly TTG composition in terms of major elements^{8,9}, but which do not have low Nb/Ta (ref. 4); these are not thought to be common in the geological record⁷.

At any one time, the thickness and make-up of the oceanic crust and its age at subduction will have varied, so that TTG were produced, but only in small volumes, in the early Archaean²¹. We conclude that the melting of garnet-amphibolite required for TTG production⁴ first became dominant in subducting slabs after about 3.2 Gyr ago. This corresponds to the first appearance of eclogite xenoliths with oceanic-crust signatures^{12–14}, and trace-element distributions demonstrate that many of these eclogites result from solid-state transformation of partially melted garnet-amphibolite⁴.

Although eclogite transformation² is largely discredited as the main driving force for subduction on the modern Earth²², it may have served to help initiate the modern style of subduction. The calculated density for the amphibole-bearing Gorgona composition increases from 3.29 to only 3.36 g cm^{−3} across the 'garnet-in' curve, and gradually to 3.53 g cm^{−3} at 1.8 GPa owing to increase of the modal abundance ratio of garnet to amphibole. Large density differences (0.5–0.6 g cm^{−3}) first occur when basaltic and gabbroic rocks dominate the subducting part of the crust, initially by allowing transformation of garnet-amphibolite into eclogite by partial melting.

Thus, until the latest Archaean the contribution of the down-going oceanic crust to slab pull was minimal. Also, subducting slabs were hotter and the subduction angle flatter than today¹⁹, so deep subduction may not have existed until eclogite-bearing slabs helped to pull the near-surface slabs deeper. Calculations show that the increased importance of chemical buoyancy in young, hot slabs tends to inhibit deep penetration²³. Only after slabs cooled further,

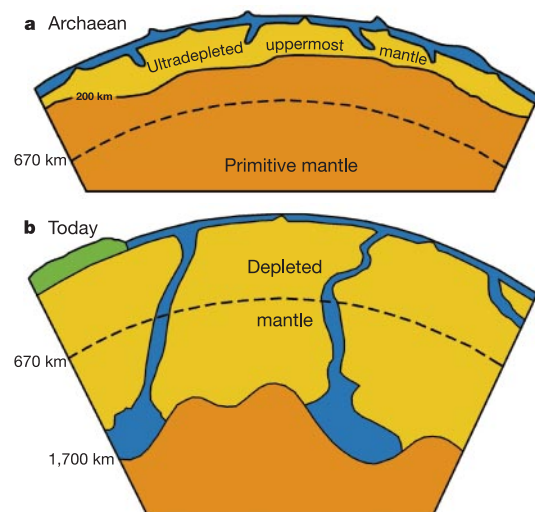


Figure 3 Model for progressive deepening of the depleted mantle with time. In contrast to the modern picture (b) where deep subduction drives the vestiges of plates into the lower mantle^{26,27}, the early Archaean oceanic crust (a) may have been produced from a layer restricted to the uppermost 200 km that consequently became strongly depleted. Only the ultramafic lower crust was returned to the mantle, and melted once again to produce more basalt. Only after thinning of the oceanic crust allowed widespread transformation of basaltic upper crust into eclogite via partial melting at the end of the Archaean may eclogite have helped to initiate the modern subduction style. Evidence for this is seen in ultradepleted trace elements in eclogite xenoliths¹³, all of which are between 3.0 and 2.6 Gyr old^{12–14}, and in fractionated Nd isotopes in the early Archaean^{5,6}.

Table 1 High-pressure experiments on the Gorgona komatiite composition

Experiment	Pressure (GPa)	Temperature (°C)	Duration (h)	Phases observed
SB-8	1.5	950	168	am + cpx + opx + gt + sp
SB-11	1.8	950	170	am + cpx + opx + gt + sp
SB-23	1.0	1,000	168	am + cpx + opx + sp
SB-26	1.2	1,000	170	am + cpx + opx + sp
SB-14	1.4	1,000	168	am + cpx + opx + gt
SB-16	1.6	1,000	169	am + cpx + opx + gt + ol
SB-15	1.8	1,000	168	am + cpx + opx + gt + ol
SB-35	1.2	1,150	144	cpx + opx + ol + sp + melt
SB-17	1.6	1,150	168	cpx + opx + ol + sp + melt
SB-18	1.8	1,150	168	cpx + opx + gt + ol + melt
SB-21	2.0	1,150	161	cpx + opx + gt + ol + melt
SB-25	2.5	1,150	168	cpx + gt + ol + melt

Abbreviations for phases: am, amphibole; cpx, clinopyroxene; opx, orthopyroxene; gt, garnet; sp, spinel; ol, olivine. The starting material for high-pressure experiments was prepared as a glass from Gorgona komatiite sample GOR-94-1, donated by C. Gignre. The composition of the glass (wt%) is SiO₂ 46.3, TiO₂ 0.66, Al₂O₃ 10.1, FeO 11.57, MgO 17.5, CaO 10.2, Na₂O 1.24, K₂O 0.03, Cr₂O₃ 0.10, Mg# 73.0. Experiments were run in a piston-cylinder apparatus with CaF₂ as pressure medium. Pt capsules with graphite inner sleeves were used, and experiments were run in water-undersaturated conditions (0.2–0.3 wt% H₂O).

and the olivine–spinel transition was reached in the underlying peridotite part of the subducting slab, could the subduction process develop into one similar to that operating today.

Evidence for this depth restriction of subduction and shallow return of crustal components to the mantle may be found in the low Nb/Ta and Zr/Hf ratios of eclogite xenoliths¹³. These are more depleted than expected for the modern Earth, and are in marked contrast to the high values presumed to represent eclogite in mixing models between depleted mantle and recycled eclogites²⁴. This apparent contradiction is resolved if the eclogite xenoliths were indeed relatively enriched for the oceanic crust of the time, but the complementary ultradepleted upper mantle was later eliminated by remixing.

Taken together, the evidence presented here supports a plate model in which only the uppermost 200 km of the mantle were tapped by oceanic volcanism in the early Archaean (Fig. 3a), as has been proposed to account for the fractionation of Nd isotopes in the early Archaean^{5,6}. This may have interesting consequences for mass-balance models between crust and mantle: geochemical models have favoured depletion of about 50% of the whole mantle to produce the crust²⁵, but are at odds with recent geophysical models for subduction into the lower mantle²⁶. Given that seismic tomography images can picture only the state of the mantle today, the geochemically modelled 50% depletion of the mantle may be a time-integrated effect of the progressive deepening of the effective depth of the mantle involved in the development of the crust. This was restricted to the uppermost 200 km in the early Archaean^{5,6}, and today involves large parts of the lower mantle²⁷ (Fig. 3b). □

Methods

Density calculations were performed by fitting linear interpolations from end-member minerals to microprobe analyses of experimental minerals, correcting for thermal expansion and compressibility²⁸, and combining with modal mineral analyses from the experiments.

Received 11 March; accepted 19 November 2002; doi:10.1038/nature01319.

1. Bickle, M. J., Nisbet, E. G. & Martin, A. Archaean greenstone belts are not oceanic crust. *J. Geol.* **102**, 121–138 (1994).
2. Green, D. H. & Ringwood, A. E. An experimental investigation of the gabbro to eclogite transformation and its petrological applications. *Geochim. Cosmochim. Acta* **31**, 767–833 (1967).
3. Schmidt, M. W. & Poli, S. Experimentally based water budgets for dehydrating slabs and consequences for arc magma generation. *Earth Planet. Sci. Lett.* **163**, 361–379 (1998).
4. Foley, F., Tiepolo, M. & Vannucci, R. Growth of early continental crust in subduction zones controlled by melting of amphibolite. *Nature* **417**, 837–840 (2002).
5. McCulloch, M. T. & Bennett, V. C. Progressive growth of the Earth's continental crust and depleted mantle: geochemical constraints. *Geochim. Cosmochim. Acta* **58**, 4717–4738 (1994).
6. Bennett, V. C., Nutman, A. P. & McCulloch, M. T. Nd isotopic evidence for transient, highly depleted mantle reservoirs in the early history of the Earth. *Earth Planet. Sci. Lett.* **119**, 299–317 (1993).
7. Martin, H. in *Archaean Crustal Evolution* (ed. Condie, K. C.) 205–259 (Elsevier, Amsterdam, 1994).
8. Winther, K. T. & Newton, R. C. Experimental melting of hydrous low-K tholeiite: evidence on the origin of Archaean cratons. *Bull. Geol. Soc. Denmark* **39**, 213–228 (1991).
9. Rapp, R. P., Watson, E. B. & Miller, C. F. Partial melting of amphibolite/eclogite and the origin of Archaean trondhjemites and tonalites. *Precamb. Res.* **51**, 1–25 (1991).
10. Zegers, T. E. & van Keken, P. E. Middle Archaean continent formation by crustal delamination. *Geology* **29**, 1083–1086 (2001).
11. de Wit, M. J. & Ashwal, L. D. (eds) *Greenstone Belts* (Oxford Univ. Press, Oxford, 1996).
12. Jacob, D., Jagoutz, E., Lowry, D., Matthey, D. & Kudryatseva, G. Diamondiferous eclogites from Siberia: remnants of Archaean oceanic crust. *Geochim. Cosmochim. Acta* **58**, 5191–5207 (1994).
13. Jacob, D. E. & Foley, S. F. Evidence for Archaean oceanic crust with low high field strength element signature from diamondiferous eclogite xenoliths. *Lithos* **48**, 317–336 (1999).
14. Shirey, S. B. et al. Archaean emplacement of eclogitic components into the lithospheric mantle during formation of the Kaapvaal craton. *Geophys. Res. Lett.* **28**, 2509–2512 (2001).
15. Farnetani, C. G., Richards, M. A. & Ghiorso, M. S. Petrological models of magma evolution and deep crustal structure beneath hotspots and flood basalt provinces. *Earth Planet. Sci. Lett.* **143**, 81–94 (1996).
16. Coogan, L. A., Thompson, G. & MacLeod, C. J. A textural and geochemical investigation of high level gabbros from the Oman ophiolite: implications for the role of the axial magma chamber at fast spreading ridges. *Lithos* **63**, 67–82 (2002).
17. Danyushevsky, L. V., Eggins, S. M., Falloon, T. J. & Christie, D. M. H₂O abundance in depleted to moderately enriched mid-ocean ridge magmas; Part I: incompatible behaviour, implications for mantle storage, and origin of regional variations. *J. Petrol.* **41**, 1329–1364 (2001).
18. Foley, S. F., Musselwhite, D. S. & van der Laan, S. R. in *Proc. VIIth Kimberlite Conf.* Vol. 1 (eds Gurney, J. J., Gurney, J. L., Pascoe, M. D. & Richardson, S. H.) 238–246 (Red Roof Design, Cape Town, 1999).
19. Karsten, J. L., Klein, E. M. & Sherman, S. B. Subduction zone geochemical characteristics in ocean ridge basalts from the southern Chile ridge: implications of modern ridge subduction systems for the Archaean. *Lithos* **37**, 143–161 (1996).

20. Abbott, D. & Mooney, W. The structural and geochemical evolution of the continental crust: support for the oceanic plateau model of continental growth. *Rev. Geophys. Suppl.* 231–242 (1995).
21. Nutman, A. P., McGregor, V. R., Friend, C. R. L., Bennett, V. C. & Kinny, P. D. The Itsaq Gneiss Complex of southern west Greenland; the world's most extensive record of early crustal evolution (3900–3600 Ma). *Precamb. Res.* **78**, 1–39 (1996).
22. Oxburgh, E. R. & Parmentier, E. M. Compositional and density stratification in oceanic lithosphere—causes and consequences. *J. Geol. Soc. Lond.* **133**, 343–355 (1977).
23. Christensen, U. R. Influence of chemical buoyancy on the dynamics of slabs in the transition zone. *J. Geophys. Res.* **102**, 22435–22443 (1997).
24. Niu, Y. & Batiza, R. Trace element evidence from seamounts for recycled oceanic crust in the Eastern Pacific mantle. *Earth Planet. Sci. Lett.* **148**, 471–483 (1997).
25. O'Nions, R. K., Evensen, N. M. & Hamilton, P. J. Differentiation and evolution of the mantle. *Phil. Trans. R. Soc. Lond. A* **297**, 479–493 (1980).
26. Widiyantoro, S. & van der Hilst, R. Structure and evolution of lithospheric slab beneath the Sunda arc, Indonesia. *Science* **271**, 1566–1570 (1996).
27. Kellogg, L. H., Hager, B. H. & van der Hilst, R. D. Compositional stratification in the deep mantle. *Science* **283**, 1881–1884 (1999).
28. Holland, T. J. B. & Powell, R. An internally consistent thermodynamic data set for phases of petrological interest. *J. Metamorph. Geol.* **16**, 309–343 (1998).
29. Liu, J., Bohlen, S. R. & Ernst, W. G. Stability of hydrous phases in subducting oceanic crust. *Earth Planet. Sci. Lett.* **143**, 161–171 (1996).
30. Wallace, M. E. & Green, D. H. The effect of bulk rock composition on the stability of amphibole in the upper mantle: implications for solidus positions and mantle metasomatism. *Mineral. Petrol.* **44**, 1–19 (1991).

Acknowledgements This Letter is dedicated to the memory of M. Rosenhauer, who prompted the initiation of the experimental programme. This work was supported by the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.F. (e-mail: sfoley@uni-greifswald.de).

Discovery of abundant hydrothermal venting on the ultraslow-spreading Gakkel ridge in the Arctic Ocean

H. N. Edmonds*, P. J. Michael†, E. T. Baker‡, D. P. Connelly§, J. E. Snow||, C. H. Langmuir¶, H. J. B. Dick#, R. Mühe☆, C. R. German§ & D. W. Graham**

* The University of Texas at Austin, Marine Science Institute, 750 Channel View Drive, Port Aransas, Texas 78373-5015, USA

† The University of Tulsa, 600 College Avenue, Tulsa, Oklahoma 74104, USA

‡ Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, 7600 Sand Point Way NE, Seattle, Washington 98115, USA

§ Southampton Oceanography Centre, Southampton SO14 3ZH, UK

|| Max-Planck Institut für Chemie, Postfach 3060, 55020 Mainz, Germany

¶ Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 02138, USA

Department of Marine Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

☆ University of Kiel, Institute of Geosciences, Olshausenstrasse 40, 24118 Kiel, Germany

** College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA

Submarine hydrothermal venting along mid-ocean ridges is an important contributor to ridge thermal structure¹, and the global distribution of such vents has implications for heat and mass fluxes² from the Earth's crust and mantle and for the biogeography of vent-endemic organisms³. Previous studies have predicted that the incidence of hydrothermal venting would be extremely low on ultraslow-spreading ridges (ridges with full spreading rates <2 cm yr⁻¹—which make up 25 per cent of the global ridge length), and that such vent systems would be hosted in ultramafic in addition to volcanic rocks^{4,5}. Here we present evidence for

active hydrothermal venting on the Gakkel ridge, which is the slowest spreading ($0.6\text{--}1.3\text{ cm yr}^{-1}$) and least explored mid-ocean ridge. On the basis of water column profiles of light scattering, temperature and manganese concentration along 1,100 km of the rift valley, we identify hydrothermal plumes dispersing from at least nine to twelve discrete vent sites. Our discovery of such abundant venting, and its apparent localization near volcanic centres, requires a reassessment of the geologic conditions that control hydrothermal circulation on ultraslow-spreading ridges.

High-temperature venting on the sea floor gives rise to plumes which ascend through the oceanic water column, entraining sea water until they reach a level of neutral buoyancy some hundreds of metres above the seabed⁶. Subsequent lateral spreading of these plumes, with their associated signatures in temperature, chemical tracers and suspended particulate material, provides a means by which new hydrothermal vent sites can be located^{2,7–9}. In 2001, the ice-breakers USCGC *Healy* and PFS *Polarstern* conducted petrological sampling and geophysical surveys along more than 1,100 km of the Gakkel ridge from 8° W to 85° E . In conjunction with this study, we conducted a reconnaissance for hydrothermal plumes

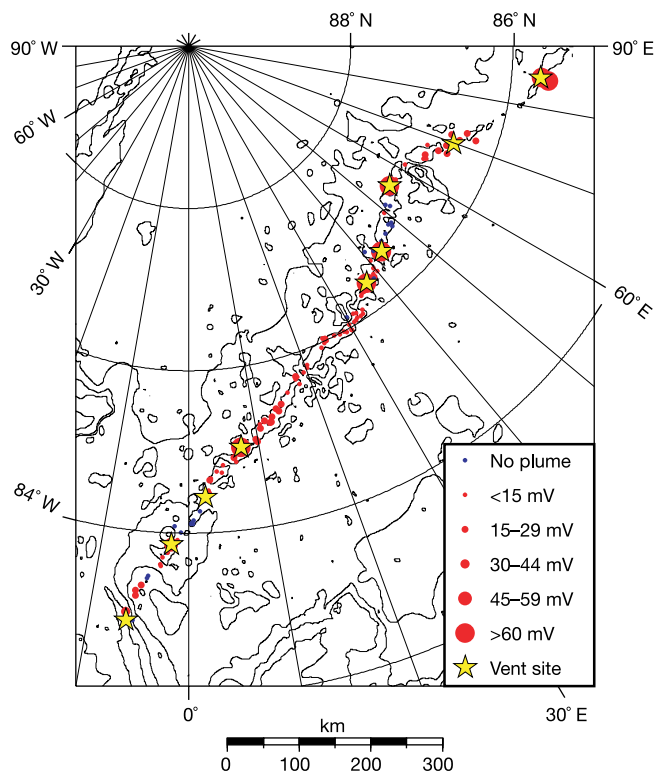


Figure 1 Map of survey area, indicating stations occupied by USCGC *Healy* and PFS *Polarstern* during the AMORE (Arctic Mid-Ocean Ridge Expedition) cruise at which MAPR profiles were obtained. (Also shown are 1,000-m depth contours from the International Bathymetric Chart of the Arctic Ocean.) MAPRs were deployed above 97 dredges, 19 wax cores, and 6 CTDs from *Healy*; and 28 TV-grabs, a heat-flow probe, and a camera tow from *Polarstern*. Symbols indicate the presence and amplitude of hydrothermal plume signals as recorded by the MAPR light scattering sensors. Optical backscatter is the primary tool by which hydrothermal plumes are detected, because temperature signals are generally extremely small and more rapidly dissipated. Increased optical backscatter results from the presence of suspended particles formed by the oxidation of hydrothermal iron and manganese on mixing with sea water. Blue circles show stations at which no plume signals were detected. Red circles are sized according to the amplitude of observed plume signals (in millivolts above background on the light-scattering sensor). Yellow stars indicate the locations of the nine hydrothermal vent sites described in Table 1.

using Miniature Autonomous Plume Recorders (MAPRs), which record temperature and light scattering (indicative of suspended particulate material) with depth¹⁰. Of 145 successful MAPR deployments, 119 (82%) showed layers of increased light scattering consistent with hydrothermal plumes, and 58 had discernible peaks in temperature correlated with the light-scattering signal (Fig. 1). This frequency of plume observations is higher than found in any previously surveyed area of the mid-ocean ridge. In three previous comparable surveys using MAPRs, the percentages of profiles showing plumes were 6–14% for the southeast Indian ridge (spreading rate $6.5\text{--}6.8\text{ cm yr}^{-1}$)⁸, 20% for the southwest Indian ridge (0.84 cm yr^{-1})¹¹, and <50% for the East Pacific Rise ($8\text{--}8.7\text{ cm yr}^{-1}$, $15^\circ 20'\text{--}18^\circ 30'\text{ N}$)¹².

The character of MAPR profiles can be used to infer proximity to the source of hydrothermal activity. Temperature anomalies associated with hydrothermal plumes are generally small ($<0.01^\circ\text{C}$) and only discernible close to the site of venting. Both the amplitude and shape of light-scattering peaks also provide information on location. Larger signals and profiles with multiple peaks indicate close proximity to the source. More than a few kilometres away from a vent site, light-scattering peaks will be smaller and smoother in shape, and temperature signals will be small if detectable. On the basis of these characteristics, we have identified nine vent sites on the Gakkel ridge to within a few kilometres (Fig. 1, Table 1). In addition to the water column data, at $6^\circ 15'\text{ W}$ (site 1) *Healy* recovered fresh sulphide chimneys in a dredge. A subsequent camera tow from *Polarstern* revealed shimmering water and abundant biological activity in this area, which we have named 'Aurora'. We also recovered hydrothermally altered rocks in several dredges,

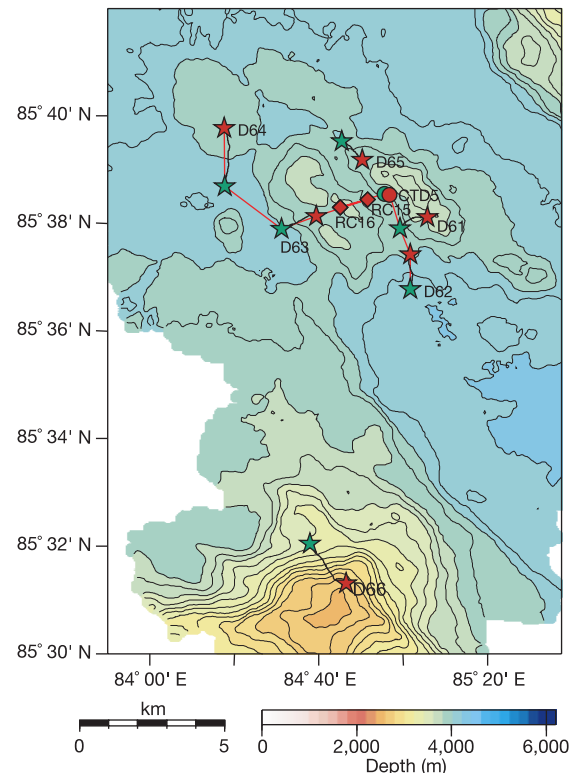


Figure 2 AMORE multibeam bathymetry of the volcanic area near 85° E , with AMORE station locations. Green and red symbols represent respectively the start and the end of a station. Stations are labelled according to the type of equipment deployed (D, dredge; RC, rock core; and CTD, conductivity–temperature–depth hydrocast). The red line indicates the path represented by the 'sections' in Fig. 3. Plumes were observed at all stations indicated on the map.

including part of the subsurface plumbing system of a relict high-temperature hydrothermal field at 30° 05' E.

The site at 85° N, 85° E (Table 1, Figs 2, 3 and 4a) is of particular interest because of its occurrence in an area of seismic^{13,14} and probably volcanic¹⁵ activity in 1999. Plumes observed in this area were extremely thick (up to 1,400 m, whereas most plumes we observed were <500 m thick), and are centred more than 1,000 m off the sea floor. Typical maximum rise height for hydrothermal plumes in the Pacific and Atlantic oceans is 200–400 m (refs 2, 6, 16). Indications from our Gakkel ridge survey suggest similar rise heights in most areas. The plumes at 85° E, in their thickness, height of rise and magnitude of signals, indicate the most vigorous hydrothermal venting seen anywhere in our survey.

Additional hydrothermal plumes were mapped that do not seem to originate from these nine vent sites (Supplementary Information). The MAPR data suggest at least three other sources: one near 2° 48' E (between sites 3 and 4) seen as a small but sharp plume in both light scattering and temperature at 2,900 m, one between 8° E and 18° E identified as a widespread plume at 3,100 to 3,200 m, and one between 21° E and 31° E identified through multiple observations of small plumes at 3,400–3,500 m. Our observations thus imply at least nine, as many as twelve, and possibly additional undiscovered vent sites along this 1,100 km section of the Gakkel ridge. These results yield a frequency of at least one active vent site per 100 km along the ridge axis.

Our results not only document evidence of active hydrothermal venting on the Gakkel ridge, but also show an unusually high percentage of MAPR profiles that have signatures consistent with hydrothermal plumes. We collected water samples at six stations in order to check that the MAPR signals do indeed indicate hydrothermal plumes; several chemical tracers such as manganese are diagnostic of a hydrothermal origin. Total dissolvable manganese (TDMn) concentrations in the 85° E plume (Fig. 4a) exceed 10 nM,

with background values <1 nM. Chemical confirmation, although important even for such large, unambiguous plumes, is most important for the many smaller MAPR signals observed on the Gakkel ridge (Fig. 1). Manganese concentrations in one such plume (Fig. 4b) clearly demonstrate the hydrothermal nature of the peak in light scattering, strengthening our interpretation of the remaining profiles. The apparent persistence of plumes on the Gakkel ridge may result from the nearly continuous nature of the axial rift valley compared with the segmentation observed on, for example, the Mid-Atlantic Ridge, which often limits the extent of an individual plume to a single ridge segment some tens of kilometres in length¹⁷.

Understanding the frequency and geologic setting of hydrothermal venting along the mid-ocean ridge system is central to estimating the fluxes of heat and matter exchanged in these systems. The Gakkel ridge is particularly important, because 25% of the total length of mid-ocean ridges spreads at <2 cm yr⁻¹ (ref. 4). Early efforts at synthesizing the global distribution of hydrothermal vents⁴ indicated a linear relationship between 'plume incidence' (defined as the percentage by length of ridge axis overlain by hydrothermal plumes) and ridge spreading rate and thus the magmatic heat input. To estimate the number of individual vent sites expected from this relationship, we assume that the plume overlying a single vent source will extend over 10 km of ridge axis². On the basis of this relationship, the length of the Gakkel ridge that we surveyed would be predicted to have 4–5 vent sites, or less than half of what we observed.

Several recent studies^{8,11,17–19} have found similar departures from the linear relationship between hydrothermal activity and spreading rate on intermediate- and slow-spreading ridges, and many have suggested that tectonic activity (faulting) is the mechanism that allows for prolonged focused venting where magmatic heat supply is low^{5,20}. Most recently, on the ultraslow-spreading portion of the southwest Indian ridge between 10° and 16° E, where there is very

Table 1 Primary vent sites identified during the AMORE cruise

Site	Location	Geologic setting	Evidence for venting
1 'Aurora'	82° 53' N, 6° 15' W	Southern flank of saddle in centre of axial rift valley	Dredge recovery of sulphide chimneys, shimmering water and abundant macrofauna observed on camera tow, light scattering peaks in MAPR data.
2	83° 51' N, 2° W	Southwestern face of volcanic edifice	MAPR: large-amplitude (50 mV) peak in light scattering with layered structure and corresponding peaks in temperature; two additional MAPR profiles in the same area showed smaller peaks at the same depth (2,400–2,600 m).
3	84° 26' N, 2° 8' E	Broad saddle in rift valley; depth of plume suggests that venting may originate on rift valley walls	Very sharp peak in MAPR light-scattering profile, 25 mV in amplitude over ~100 m depth interval (centred at 3,000 m) with corresponding temperature shift of ~0.006 °C.
4	85° 1' N, 7° 27' E	Elongate ridge on northwestern side of rift valley	MAPR profile with 86 mV light-scattering peak between 2,700 and 2,900 m, with an associated temperature peak of ~0.01 °C above background. Note dredging on this ridge recovered only peridotite.
5	86° 21' N, 37° E	Volcanic ridge, slightly north of centre of rift valley	MAPR: multiple observations of light-scattering peaks at ~3,200 m, up to 70 mV in amplitude, with temperature peaks corresponding to the largest LSS peaks. A CTD cast provided confirming evidence through hydrography, transmissometry and manganese concentrations. Peaks in multiple profiles suggest at least two sources of venting on this ridge.
6	86° 32' N, 43° E	Peak of volcanic mound on southern side of axial rift valley	Large, multilayered MAPR peaks in light scattering (up to 110 mV) and temperature (~0.015 °C); confirming hydrographic and manganese data from a CTD cast.
7	86° 59' N, 55° 30' E	Northwest side of axial volcanic ridge	MAPR: large multilayered peak in light scattering (maximum anomaly > 100 mV at ~2,600 m), with corresponding temperature peaks.
8	86° 32' N, ~69° E	Axial volcanic mound	MAPR: multiple observations of light scattering (up to 30 mV) and temperature peaks at ~2,400 m.
9	85° 39' N, 84° 50' E	Axial volcanic mound	MAPR: multiple observations of large (60 mV LSS, 0.05 °C), extremely thick (up to 1,400 m) plumes centred at ~2,500 m; additional plume layers at ~3,300 m suggestive of second site; confirming hydrographic and manganese data from one CTD cast.

Supporting documentation for each site is available as Supplementary Information. LSS, light-scattering sensor; CTD, conductivity–temperature–depth profiler.

little volcanic activity, several hydrothermal sites were discovered associated with large faults and hosted in ultramafic rocks¹¹. Even the fast-spreading northern East Pacific Rise, where hydrothermal activity would be expected to be most closely tied to the magma budget, has been found to have a wide range of vent frequencies at similar spreading rates¹², indicating that on the scale of individual ridge segments the distribution of vents is controlled by the degree of faulting in the volcanic rocks. Extending these results to the Gakkel ridge, we expected to find most of our vent sites in highly faulted regions and predominantly associated with exposures of mantle rocks. Chemical differences between volcanic and ultramafic-hosted systems (and their effect on global ridge fluxes) have been highlighted in several studies of hydrothermal venting on slow-spreading ridges^{11,21–23}. However, eight of the nine sites we have described are found in association with volcanic constructional features rather than with clearly tectonic features such as the rift valley walls. Only one (site 4) seems to be associated with ultramafic rocks. Our results indicate that this style of venting may not be as ubiquitous as recently presumed, and may require a new model for vent distribution at ultraslow spreading rates. Specifically, why are the vents on the Gakkel ridge localized at volcanoes, unlike on other slow-spreading ridges?

The basic requirements for hydrothermal circulation are a source of heat and a pathway for sea water to circulate within the rock. Slower spreading rates imply a weaker thermal gradient (less heat input at the ridge axis)¹, and thus deep-thrusting faults are required in order for circulating sea water to reach high temperatures on slow-spreading ridges. On the Mid-Atlantic Ridge and the south-west Indian ridge such faults enable the establishment of high-

temperature vent systems, often in areas quite distant from ridge volcanoes^{5,11}. We suggest that on the ultraslow-spreading Gakkel ridge, faulting cannot be the dominant control on the location of high-temperature venting, because of the overall thermally cool character of the crust^{24,25}. Although faulting is still required, the implication is that the volcanic centres, which are both localized and widely spaced on the Gakkel ridge²⁶, are the only areas of the ridge with enough heat to drive high-temperature convection and give rise to the plumes we observed. Verification of this suggestion, and a fuller understanding of the distribution of volcanic^{14,15} and hydrothermal activity on the Gakkel ridge, will require location and

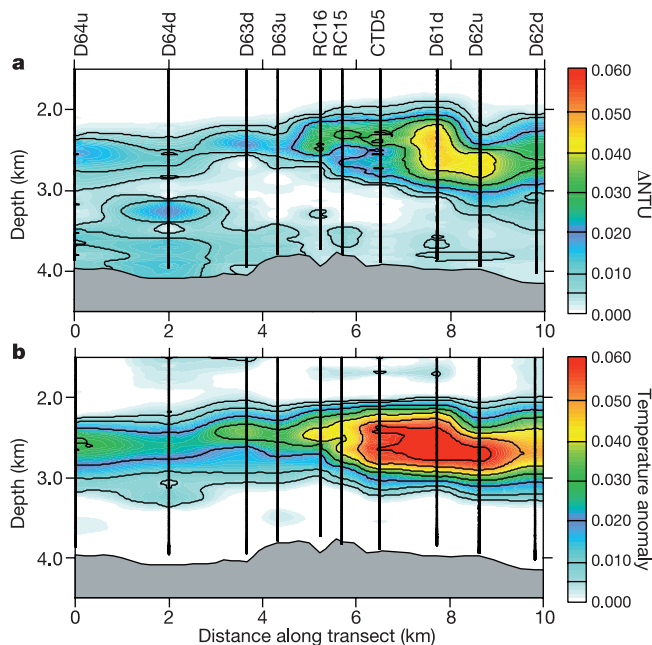


Figure 3 Light scattering and temperature data from the 85°E area, along the track indicated in Fig. 2. Vertical lines indicate the location of the profiles used for the section. **a**, Light scattering anomalies in NTU (nephelometric turbidity units, calculated from instrument voltage and defined in ref. 12). Δ NTU is the difference between the NTU value at a particular depth and that for the profile background. Yellows and reds thus indicate the strongest plume signals. Layering of the plume is most pronounced over the centre of the volcano near RC15 and CTD5, whereas the plume is thickest and the amplitude strongest near D61. **b**, Temperature anomalies (in °C) along the same section, defined as the difference in temperature between the measured profile and a polynomial fit to the background temperature profile.

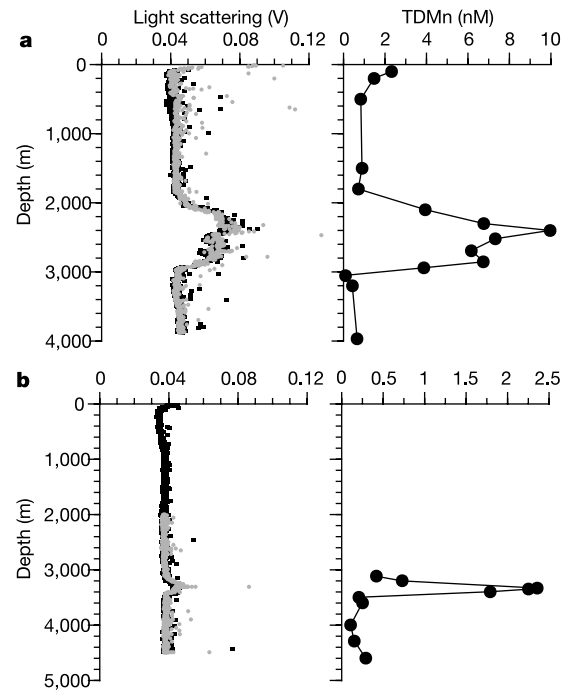


Figure 4 Light-scattering sensor and total dissolvable manganese (TDMn) data from AMORE CTD stations 5 and 9. The concentration of dissolved manganese is typically enriched by a factor of over a million in hydrothermal vent fluids compared to deep sea water, and is therefore still enriched even at the high dilution factors ($\sim 10,000:1$) characteristic of hydrothermal plumes⁷. We measured TDMn instead of filtering the samples to determine solely dissolved Mn; therefore, some portion of the signal could result from leaching of manganese oxide coatings off suspended particles. Our TDMn concentrations, however, are comparable to those observed in other hydrothermal plumes, and higher than would be expected for acid-leachable Mn from the concentrations of suspended particulate material that are indicated by the light-scattering profiles^{2,7,30}, indicating along with the temperature data that the peaks in light scattering do not result from resuspension of bottom sediments off the rift valley walls and into the mid-water column of the axial valley. **a**, Light-scattering and manganese data for CTD 5 at 86°N, 85°E (Fig. 2). Similar results were obtained at sites 5 and 6 (Table 1 and Supplementary Information). Hydrocasts near sites 2 and 7 failed to intercept the plumes we had observed with the MAPRs, owing to difficulty positioning the ship in the ice. Note that for MAPR light-scattering profiles (including those found in the Supplementary Information), black symbols indicate the downcast profile of the station, and grey symbols are from the upcast. Higher values in volts from the light-scattering sensor indicate greater concentrations of suspended particulate matter in the water. Water samples for TDMn determination were collected on the upcast. **b**, Light-scattering and manganese data from CTD9 at 85°59'N, 24°42'E. This station was designed to sample one of the ubiquitous smaller plumes observed with the MAPR light-scattering sensors (Fig. 1), in order to verify with chemical tracers that the physical signals observed were in fact hydrothermal in origin. The clear peak in manganese concentration coincident with the increase in suspended particulate matter provides confirmation of this and, by inference, the dozens of other MAPR light-scattering peaks observed.

detailed surveys of hydrothermal sites and additional geophysical studies.

One of the factors driving the continuing search for hydrothermal vents on the Gakkel ridge is the unknown make-up of the chemo-synthetic faunal communities that surround them. The fauna inhabiting known Atlantic and Pacific vent sites are markedly different from each other, and also exhibit distinct biogeographical differences within each of these oceans^{3,27}. There is no modern deep-water (ridge depth) connection between the Gakkel ridge and other parts of the mid-ocean ridge system south of Iceland. Current knowledge of the tectonic history of the Arctic Ocean^{28,29} suggests that there has been no such connection between the Arctic and the other major ocean basins during its history. Therefore it is likely that new species of vent biota, which have evolved in isolation from those in other oceans, will be discovered at hydrothermal sites on the Gakkel ridge³. Future studies that locate these vent sites, and return images and samples from them, will provide insights into the evolution of vent organisms and the pathways by which they migrated into these sites. □

Received 1 October; accepted 26 November 2002; doi:10.1038/nature01351.

- Chen, Y. J. in *Ophiolites and Oceanic Crust: New Insights from Field Studies and the Ocean Drilling Program* (eds Dilek, Y., Moores, E. M., Elthon, D. & Nicolas, A.) 161–179 (Geological Society of America, Boulder, Colorado, 2000).
- Baker, E. T., German, C. R. & Elderfield, H. in *Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions* (eds Humphris, S. E., Zierenberg, R. A., Mullineaux, L. S. & Thomson, R. E.) 47–71 (American Geophysical Union, Washington DC, 1995).
- Van Dover, C. L., German, C. R., Speer, K. G., Parson, L. M. & Vrijenhoek, R. C. Evolution and biogeography of deep-sea vent and seep invertebrates. *Science* **295**, 1253–1257 (2002).
- Baker, E. T., Chen, Y. J. & Phipps Morgan, J. The relationship between near-axis hydrothermal cooling and the spreading rate of mid-ocean ridges. *Earth Planet. Sci. Lett.* **142**, 137–145 (1996).
- German, C. R. & Parson, L. M. Distributions of hydrothermal activity along the Mid-Atlantic Ridge: interplay of magmatic and tectonic controls. *Earth Planet. Sci. Lett.* **160**, 327–341 (1998).
- Lupton, J. E., Delaney, J. R., Johnson, H. P. & Tivey, M. K. Entrainment and vertical transport of deep-ocean water by buoyant hydrothermal plumes. *Nature* **316**, 621–623 (1985).
- Klinkhammer, G., Rona, P., Greaves, M. & Elderfield, H. Hydrothermal manganese plumes in the Mid-Atlantic Ridge rift valley. *Nature* **314**, 727–731 (1985).
- Scheirer, D. S., Baker, E. T. & Johnson, K. T. M. Detection of hydrothermal plumes along the Southeast Indian Ridge near the Amsterdam–St. Paul Plateau. *Geophys. Res. Lett.* **25**, 97–100 (1998).
- Charlou, J. L. & Donval, J. P. Hydrothermal methane venting between 12°N and 26°N along the Mid-Atlantic Ridge. *J. Geophys. Res.* **98**, 9625–9642 (1993).
- Baker, E. T. & Milburn, H. B. MAPR: A new instrument for hydrothermal plume mapping. *Ridge Events* **8**, 23–25 (1997).
- Bach, W., Banerjee, N. R., Dick, H. J. B. & Baker, E. T. Discovery of ancient and active hydrothermal systems along the ultra-slow spreading Southwest Indian Ridge 10°–16°E. *Geochim. Geophys. Geosyst.* **3**, 101029/2001GC000279 (2002).
- Baker, E. T., Cormier, M.-H., Langmuir, C. H. & Zavala, K. Hydrothermal plumes along segments of contrasting magmatic influence, 15°20'N–18°30'N, East Pacific Rise: Influence of axial faulting. *Geochim. Geophys. Geosyst.* **2**, 2000GC000165 (2001).
- Müller, C. & Jokat, W. Seismic evidence for volcanic activity discovered in central Arctic. *Eos* **81**, 265–269 (2000).
- Tolstoy, M., Bohnenstiehl, D. R., Edwards, M. H. & Kurras, G. J. Seismic character of volcanic activity at the ultraslow-spreading Gakkel Ridge. *Geology* **29**, 1139–1142 (2001).
- Edwards, M. H. *et al.* Evidence of recent volcanic activity on the ultraslow-spreading Gakkel ridge. *Nature* **409**, 808–812 (2001).
- Speer, K. G. & Rona, P. A. A model of an Atlantic and Pacific hydrothermal plume. *J. Geophys. Res.* **94**, 6213–6220 (1989).
- German, C. R., Parson, L. M. & the HEAT Scientific Team. Hydrothermal exploration near the Azores Triple Junction: tectonic control of venting at slow-spreading ridges? *Earth Planet. Sci. Lett.* **138**, 93–104 (1996).
- German, C. R., Baker, E. T., Mevel, C., Tamaki, K. & the FUJI Scientific Team. Hydrothermal activity along the southwest Indian Ridge. *Nature* **395**, 490–493 (1998).
- German, C. R. *et al.* Hydrothermal activity on the Reykjanes Ridge: the Steinahóll vent-field at 63°06'N. *Earth Planet. Sci. Lett.* **121**, 647–654 (1994).
- Kleinrock, M. C. & Humphris, S. E. Structural control on sea-floor hydrothermal activity at the TAG active mound. *Nature* **382**, 149–153 (1996).
- Fouquet, Y. *et al.* FLORES diving cruise with the Nautilie near the Azores—First dives on the Rainbow field: hydrothermal seawater/mantle interaction. *InterRidge News* **7**, 24–28 (1998).
- Charlou, J. L. *et al.* Intense CH₄ plumes generated by serpentinization of ultramafic rocks at the intersection of the 15°20'N fracture zone and the Mid-Atlantic Ridge. *Geochim. Cosmochim. Acta* **62**, 2323–2333 (1998).
- Kelley, D. S. *et al.* An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30°N. *Nature* **412**, 145–149 (2001).
- Coakley, B. J. & Cochran, J. R. Gravity evidence of very thin crust at the Gakkel Ridge (Arctic Ocean). *Earth Planet. Sci. Lett.* **162**, 81–95 (1998).
- Bown, J. W. & White, R. S. Variation with spreading rate of oceanic crustal thickness and geochemistry. *Earth Planet. Sci. Lett.* **121**, 435–449 (1994).
- Kurras, G. J. *et al.* Axial valley morphology of the Gakkel Ridge [8°W–88°E]: Seabeam and Hydrosweep bathymetry from the Arctic Mid-Ocean Ridge Expedition (AMORE 2001). *Eos* **82**

- (Fall Meeting Suppl.), T11B-0855 (2001).
- Tunnicliffe, V. & Fowler, C. M. R. Influence of sea-floor spreading on the global hydrothermal vent fauna. *Nature* **378**, 531–533 (1996).
- Johnson, G. L., Pogrebitsky, J. & Macnab, R. in *The Polar Oceans and Their Role in Shaping the Global Environment* (eds Johannessen, R. E., Muench, R. D. & Overland, J. E.) 285–294 (American Geophysical Union, Washington DC, 1994).
- Lawver, L. A., Muller, R. D., Srivastava, S. P. & Roest, W. in *Geological History of the Polar Oceans: Arctic versus Antarctica* (eds Bleil, U. & Thiede, J.) 29–62 (Kluwer Academic, Dordrecht, 1990).
- Baker, E. T., Tennant, D. A., Feely, R. A., Lebon, G. T. & Walker, S. L. Field and laboratory studies on the effect of particle size and composition on optical backscattering measurements in hydrothermal plumes. *Deep-Sea Res.* **1** **48**, 593–604 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank all the members of the AMORE science party for their assistance, and the officers and crews of the USCGC *Healy* and PFS *Polarstern* for their logistical and technical support. We also thank S. Walker for assistance with MAPR support and Fig. 3. This work was supported by the US National Science Foundation, the US NOAA Vents Program, a Heisenberg fellowship to J.E.S. from the Deutsche Forschungsgemeinschaft, and W. S. Gardner and the University of Texas at Austin.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.N.E. (e-mail: edmonds@utmsi.utexas.edu).

Increased CO₂ uncouples growth from isoprene emission in an agriforest ecosystem

Todd N. Rosenstiel^{†‡}, Mark J. Potosnak^{†§||}, Kevin L. Griffin[§], Ray Fall^{‡¶} & Russell K. Monson^{‡‡}

^{*} Department of Environmental, Population, and Organismic Biology,

[‡] Cooperative Institute for Research in Environmental Sciences,

[¶] Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309, USA

[§] Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA

[†] These authors contributed equally to this work

The emission of isoprene from the leaves of forest trees is a fundamental component of biosphere–atmosphere interactions, controlling many aspects of photochemistry in the lower atmosphere^{1–3}. As almost all commercial agriforest species emit high levels of isoprene⁴, proliferation of agriforest plantations has significant potential to increase regional ozone pollution^{5–7} and enhance the lifetime of methane⁸, an important determinant of global climate. Here we show that growth of an intact *Populus deltoides* plantation under increased CO₂ (800 μmol mol^{−1} and 1,200 μmol mol^{−1}) reduced ecosystem isoprene production by 21% and 41%, while above-ground biomass accumulation was enhanced by 60% and 82%, respectively. Exposure to increased CO₂ significantly reduced the cellular content of dimethylallyl diphosphate, the substrate for isoprene synthesis, in both leaves and leaf protoplasts. We identify intracellular metabolic competition for phosphoenolpyruvate as a possible control point in explaining the suppression of isoprene emission under increased CO₂. Our results highlight the potential for uncoupling isoprene emission from biomass accumulation in an agriforest species, and show that negative air-quality effects of proliferating agriforests may be offset by increases in CO₂.

|| Present address: National Center for Atmospheric Research, Boulder, Colorado 80305, USA.

Unlike other crops, almost all fast-growing agriforest species, including *Acacia* (tropical), *Eucalyptus* (tropical) and *Populus* (temperate), emit significant quantities of isoprene (2-methyl-1,3-butadiene)^{4,9}, a highly reactive hydrocarbon. Afforestation with such fast-growing trees has been suggested as a way to ameliorate increases in atmospheric CO₂ concentration¹⁰, [CO₂], and plantation forests will undoubtedly be important in supplying future global demand for wood¹¹. When such plantations occur on large spatial scales, the effects on atmospheric chemistry can be marked¹². As a result, the continued expansion of agriforest plantations (10.5 Mha yr⁻¹ at present)¹³ has the potential to affect significantly the oxidative behaviour of the atmosphere, in turn negatively affecting local air quality^{5,6}, perturbing regional biogeochemical cycles³ and further contributing to climate forcing by enhancing the lifetime of tropospheric methane⁸. However, realistic estimates of regional isoprene emission are difficult to obtain, and we also lack an adequate understanding of the biochemical mechanisms that may couple isoprene emission to increased [CO₂] and associated global climate change.

To examine the effect of [CO₂] on isoprene emission of an agriforest species, we established three cottonwood plantations (*Populus deltoides* Bartr.), from approximately 50 trees each, in the forestry section of Columbia University's Biosphere 2 Center, a large controlled-environment research facility¹⁴. The poplar clone used was obtained from Westvaco Corporation and is currently used as planting stock for *Populus* agriforests in the south-eastern United States. Poplar clones were planted in 1998 in three large replicate mesocosms. [CO₂] set points within the three mesocosms have been maintained at 430 μmol mol⁻¹ (p.p.m.), 800 p.p.m. and 1,200 p.p.m. since May 2000. By the end of the growing season (day 341), canopy height ranged from 7.5–8.1 m (approximately one-half mesocosm height).

The semi-closed nature of the Biosphere 2 facility (see Methods) allowed a continuous determination of net ecosystem isoprene production. Figure 1 shows the typical day-to-day variability and pattern of isoprene emission in the three mesocosms. All three mesocosms displayed the well-characterized light-dependent diurnal pattern of isoprene emission¹⁵, with maximal emissions occurring at roughly midday. Compared with the current [CO₂] of 430 p.p.m., growth at increased [CO₂] resulted in marked decreases in the daily release of isoprene (Fig. 1). Here we report the effects of atmospheric [CO₂] on whole-ecosystem isoprene emissions; the results are similar to previous reports of leaf-level suppression of isoprene emission in *Populus*¹⁶ in response to increased [CO₂].

When integrated over the 2000 growing season, net ecosystem isoprene production was reduced at increased [CO₂] by 21% and

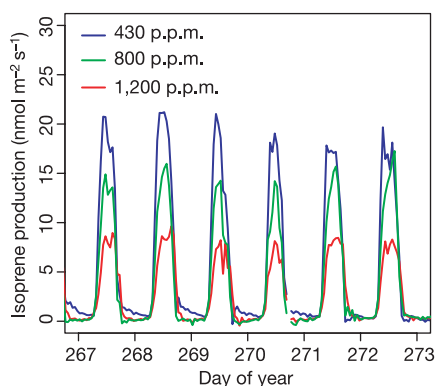


Figure 1 Net ecosystem isoprene production for *Populus deltoides* plantations. Data is shown for growth at three different concentrations of CO₂ at the Biosphere 2 facility. The figure depicts six representative days during the 2000 growing season. Isoprene production is expressed per ecosystem area.

41%, respectively (Table 1). However, reductions in ecosystem isoprene production were not a consequence of decreased productivity. Total above-ground plant biomass harvested at the end of the growing season increased by 60% and 82% with increased [CO₂] (Table 1). We also measured leaf-level photosynthesis rates throughout the growing season, which were generally enhanced under increased [CO₂] (data not shown). Throughout the course of the experiment, the poplar leaves displayed no indication of photosynthetic downregulation at either 800 or 1,200 p.p.m. CO₂. Instead, the biomass data suggests a significant degree of CO₂ fertilization of this rapidly growing species, supporting numerous other studies suggesting that the productivity and capacity to sequester carbon in *Populus* sp. may benefit from enhanced [CO₂]¹⁷. To summarize the mesocosm results, it appears that the percentage of fixed CO₂ converted to isoprene is decreased by CO₂ fertilization.

To study the biochemical nature of the isoprene suppression, we measured the content of dimethylallyl diphosphate (DMAPP) in leaves growing in each of the three mesocosms. DMAPP is the immediate precursor of isoprene biosynthesis, and the regulation of its production in plant chloroplasts has been implicated in the control of isoprene emission¹⁸. Isoprene-emission rates were determined on leaves ($n = 6-12$) from each mesocosm in the middle of the photoperiod and subsequently harvested for DMAPP analysis. Both isoprene-emission rates (normalized to leaf area) and DMAPP content were significantly reduced at increased [CO₂], and there was a significant positive correlation between DMAPP content and

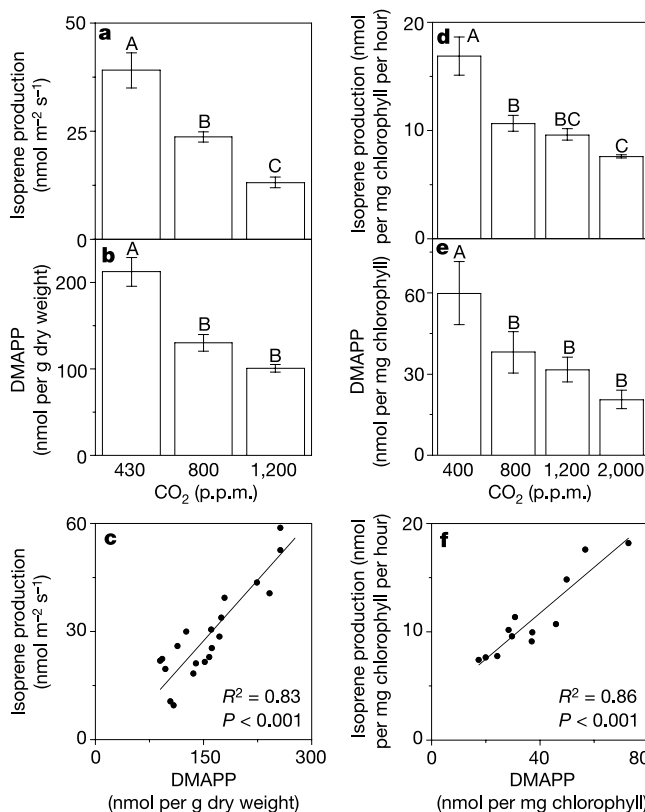


Figure 2 Isoprene production and dimethylallyl diphosphate (DMAPP) content in leaves and protoplasts of *P. deltoides*. **a-c**, Isoprene production (**a**), DMAPP content (**b**) and their relationship with a fitted linear regression (**c**) from leaves collected at the Biosphere 2 facility. Values represent the mean $\pm$ s.e.m. ($n = 6-8$). **d-f** Isoprene production (**d**), DMAPP content (**e**) and their relationship with a fitted linear regression (**f**) from protoplasts incubated under four concentrations of CO₂. Values represent the mean $\pm$ s.e.m. ($n = 3$). Within a panel, means with the same letter are not significantly different ($P < 0.05$).

Table 1 Isoprene production and biomass accumulation at varying [CO₂]

[CO ₂] (p.p.m.)	Isoprene production (mmol m ⁻²)	Biomass (g m ⁻²)
430	79.5	668
800	61.4	1,070
1,200	46.9	1,220

Values for the net ecosystem isoprene production and the above-ground biomass accumulation are expressed per ecosystem area. Measures were obtained for three *Populus deltoides* plantations grown at three concentrations of CO₂ ([CO₂]) in the Biosphere 2 facility during the 2000 growing season.

isoprene emission (Fig. 2a–c). These data indicate that controls over DMAPP synthesis underlie the suppression of isoprene emission at increased [CO₂].

To further examine this possibility, we developed an independent experimental system using isolated protoplasts. Intact, isoprene-emitting protoplasts were isolated from leaves of the same clones of *P. deltoides* Bartr. growing in greenhouses at the University of Colorado, Boulder. When incubated under increased [CO₂], protoplasts showed the same suppression of isoprene production and DMAPP content (Fig. 2d–f), and the same strong positive correlation between DMAPP content and isoprene production as leaves grown under increased [CO₂]. Isoprene emission from protoplasts was found to be especially sensitive to inhibition by NaHCO₃ (Fig. 3), an alternative CO₂ source. Once again, the suppression of isoprene emission was strongly positively correlated with the cellular concentration of DMAPP. Taken together, results from both whole leaves and isolated protoplasts strongly support the hypothesis that the suppression of isoprene emission under increased [CO₂] involves substrate-level regulation and is a direct result of reduced DMAPP production.

The controls over DMAPP synthesis in chloroplasts are poorly understood¹⁹. However, recent results indicate that a significant fraction of carbon entering the chloroplastic DMAPP pool is extra-chloroplastic in origin²⁰, and may arise from cytosolic phosphoenolpyruvate (PEP). If so, then competition for PEP in the plant cell cytosol could have an impact on isoprene emission in the chloroplast. We proposed that increasing cellular [HCO₃⁻], in response to increasing [CO₂], might directly reduce cytosolic PEP concentrations by stimulating the activity of cytosolic PEP-carboxylase (PEPC). PEPC, a ubiquitous plant enzyme, catalyses the carboxylation of PEP to oxaloacetic acid for anaplerotic reactions in plant cells²¹, and is present with high activity in the leaves of *Populus* (data not shown). To test the role of PEPC in influencing isoprene formation, we fed two PEPC inhibitors to excised leaves of *P. deltoides* (Fig. 4). Addition of either diethyloxalacetate (DOA) or 3,3-dichloro-2-(dihydroxyphosphinoylmethyl)-propenoate (DCDP, see Methods) enhanced the rates of isoprene emission,

supporting the hypothesis that isoprene emission is responsive to the availability of cytosolic PEP. These general results were replicated several times, and experiments with *Eucalyptus* and northern red oak, both isoprene emitters, yielded similar results (data not shown). The increase in isoprene emission occurred despite variable impacts on photosynthesis. When PEPC was assayed in whole-leaf extracts, addition of either DOA or DCDP led to a significant inhibition of activity (data not shown).

It should be noted that growth in increased [CO₂] has been shown to increase both leaf mitochondrial number²² and the rate of leaf respiration in the light²³. Presumably, increased conversion of PEP to pyruvate must occur under increased [CO₂] to provide the necessary substrate for increased mitochondrial respiration during the day, further reducing the availability of cytosolic PEP for other cellular processes. We speculate that suppression of isoprene emission reflects a more fundamental metabolic response of plants to increases in [CO₂], namely, altered partitioning of PEP between mitochondrial and chloroplastic processes. In addition, this dynamic exchange between the mitochondria and the chloroplast may help to explain why the CO₂ effect on isoprene emission has been variable¹⁶.

The emission of isoprene from agriforest species represents a significant unintended consequence of agriforest cultivation. As short-rotation agriforests provide an increasing array of human services—including lumber, pulp and paper production, biofuels for renewable energy and, potentially, carbon sequestration^{10,11}—the continued widespread establishment of these managed ecosystems is likely to influence the continued rise in tropospheric ozone^{7,8} as well as the sustainability of the oxidative power of the atmosphere²⁴. Although isoprene emission is generally considered to be closely linked to photosynthetic metabolism²⁵, results from this study show that leaf development under increased [CO₂] reduced isoprene emission, despite concomitant increases in both photosynthesis and biomass accumulation. The identification of intracellular metabolic competition for PEP provides a mechanism for reconciling the decreased chloroplastic isoprene emission with the, presumably, increased rates of mitochondrial respiration necessary for biomass accumulation. The results from this study confirm

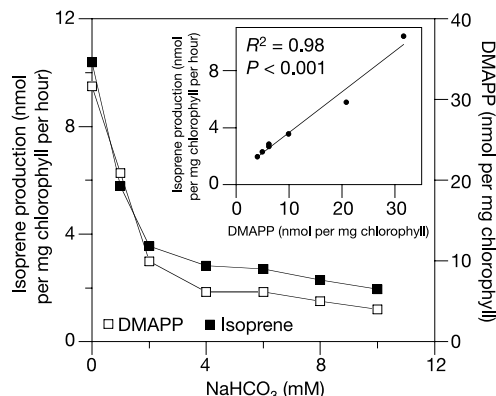


Figure 3 NaHCO₃ reduces isoprene production and DMAPP content in *P. deltoides* protoplasts. Inset, the relationship between isoprene production and DMAPP content with a fitted linear regression.

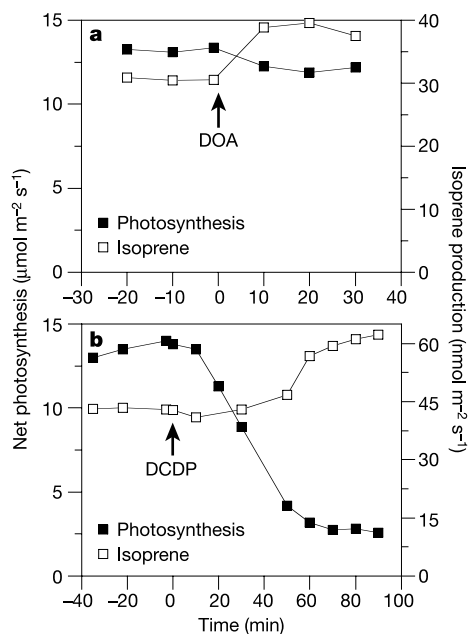


Figure 4 Influence of two inhibitors of phosphoenolpyruvate carboxylase activity on isoprene production and net photosynthesis in excised leaves of *P. deltoides*. **a**, Diethyloxalacetate (DOA; 100 μM) **b**, 3,3-dichloro-2-(dihydroxyphosphinoylmethyl)-propenoate (DCDP; 4 mM). Arrows and time zero indicate the application of inhibitor.

experimentally in an intact *Populus* ecosystem that increased [CO₂] cannot only enhance biomass accumulation in a short-rotation agriforest, but also reduce net ecosystem isoprene production. So the negative air-quality impacts (that is, isoprene emission) of proliferating agriforests may be partially offset by increases in global [CO₂]. □

Methods

Biosphere 2 experiment

This experiment was conducted in the 2,000 m² forestry section of Columbia University's Biosphere 2 Center (Oracle, Arizona, USA)¹⁴. The forestry section is physically isolated from the remainder of Biosphere 2, and has independent climate and CO₂ control; it is subdivided into three roughly equal 12,000 m³ mesocosms. Outside air is constantly exchanged with a set of push-pull fans, and [CO₂] within each bay is maintained at set points by adding pure CO₂ from a liquid storage tank. The mesocosms are subjected to the light regimes of a temperate desert region, attenuated by the glass and steel structure. An automated climate-control system maintains temperature, dew point and soil moisture (for further details, see ref. 26). Cottonwood (*P. deltoides* Bartr.) clones (S7c8-day neutral) used in this experiment were obtained from a fibre production farm (Westvaco Corporation); they are adapted to the lower Brazos River, Texas. Clones were planted in 1998 and coppiced at the end of each growing season. Each mesocosm contained 43–47 trees. We calculated hourly ecosystem isoprene production from mesocosm isoprene concentrations measured with a Fast Isoprene Sensor (Hills Scientific); volumetric exchange rates and leak rates were determined from SF6 injection, using ten-day averages to fill data voids.

Leaf photosynthesis and isoprene measurements

Leaf-level gas-exchange measurements were made with a portable photosynthesis system (Li-6400, Li-Cor Inc.). All measurements were made on attached leaves at 30 °C and 1,000 μmol photons m⁻² s⁻¹, and with growth [CO₂] at 430 p.p.m., 800 p.p.m. or 1,200 p.p.m. within Biosphere 2; in the case of the inhibitor experiments, [CO₂] was at 360 p.p.m. for detached leaves from trees growing at ambient [CO₂] in greenhouses at the University of Colorado. To determine leaf-level isoprene-emission rates, a small sample of air exiting the cuvette was injected into a gas chromatograph and measured with either photo-ionization or reduction gas detection¹⁵.

Protoplast experiments

Protoplasts were isolated from young leaves of *P. deltoides* Bartr. as described in ref. 27, with slight modification to enhance yield²⁸. To determine isoprene production, protoplasts, equivalent to 25 μg of chlorophyll, were incubated in small vials containing 0.7 M sorbitol, 50 mM HEPES (pH = 7.0), 2 mM MgCl₂ and 1 mM CaCl₂, and head-space isoprene concentrations were determined by gas chromatography (reduction gas detection) after a 30–60-min incubation at 1,000 μmol m⁻² s⁻¹ at 25 °C. CO₂ was provided as either CO₂ gas occupying the vial headspace or as NaHCO₃.

DMAPP analysis

DMAPP content was determined as isoprene released following acidification of lyophilized leaf material¹⁸, or after rapid acidification of intact protoplasts and subsequent head-space analysis.

PEPC inhibitors

Stock solutions of DOA and DCDP were prepared as described^{29,30} and supplied to the cut petiole (final concentrations: 100 μM DOA and 4 mM DCDP).

Received 14 October; accepted 15 November 2002; doi:10.1038/nature01312.
Published online 5 January 2003.

1. Fehsenfeld, F. C. *et al.* Emissions of volatile organic compounds from vegetation and the implications for atmospheric chemistry. *Glob. Biogeochem. Cycles* **6**, 389–430 (1992).
2. Fuentes, J. D. *et al.* Biogenic hydrocarbons in the atmospheric boundary layer: a review. *Bull. Am. Meteorol. Soc.* **81**, 1537–1575 (2000).
3. Monson, R. K. & Holland, E. A. Biospheric trace gas fluxes and their control over tropospheric chemistry. *Annu. Rev. Ecol. Syst.* **32**, 547–576 (2001).
4. Harley, P. C., Monson, R. K. & Lerdau, M. T. Ecological and evolutionary aspects of isoprene emission from plants. *Oecologia* **118**, 109–123 (1999).
5. Trainer, M. *et al.* Models and observations of the impact of natural hydrocarbons on rural ozone. *Nature* **329**, 705–707 (1987).
6. Chameides, W. L., Lindsay, R. W., Richardson, J. & Kiang, C. S. The role of biogenic hydrocarbons in urban photochemical smog: Atlanta as a case study. *Science* **241**, 1473–1475 (1988).
7. Pierce, T. *et al.* Influence of increased isoprene emissions on regional ozone modeling. *J. Geophys. Res.* **103**, 25611–25629 (1999).
8. Poisson, N., Kanakidou, M. & Crutzen, P. J. Impact of non-methane hydrocarbons on tropospheric chemistry and the oxidizing power of the global troposphere: three-dimensional modeling results. *J. Atmos. Chem.* **36**, 157–230 (2000).
9. Lerdau, M. & Slobodkin, L. Trace gas emissions and species-dependent ecosystem services. *Trends Ecol. Evol.* **17**, 309–312 (2002).
10. Brown, S., Sathaye, J., Cannell, M. & Kauppi, P. in *Climate Change in 1995—Impacts, Adaptations and Mitigation of Climate Change: Scientific–Technical Analyses. Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change* (eds Watson, R. T., Zinowera, M. C. & Moss, R. H.) 776–797 (Cambridge Univ. Press, Cambridge, 1996).
11. Fenning, T. M. & Gershenson, J. Where will the wood come from? Plantation forests and the role of

biotechnology. *Trends Biotechnol.* **20**, 291–296 (2002).

12. Muller, J. F. Geographical distribution and seasonal variation of surface emissions and deposition velocities of atmospheric trace gases. *J. Geophys. Res.* **97**, 3787–3804 (1992).
13. *Forest Products 1991–1995*. FAO Forestry Series No. 30, FAO Statistic Series No. 137/(Food and Agriculture Organization of the United Nations, Rome, 1995).
14. Marino, B. D. V. & Odum, H. T. *Biosphere 2 Research Past and Present* (Elsevier, New York, 1999).
15. Harley, P., Guenther, A. & Zimmerman, P. Environmental controls over isoprene emission in deciduous oak canopies. *Tree Physiol.* **17**, 705–714 (1997).
16. Sharkey, T. D., Loreto, F. & Delwiche, F. High carbon dioxide and sun/shade effects on isoprene emission from oak and aspen tree leaves. *Plant/Cell Environ.* **14**, 333–338 (1991).
17. Gielen, B. & Ceulemans, R. The likely impact of rising atmospheric CO₂ on natural and managed *Populus*: a literature review. *Environ. Pollut.* **115**, 335–338 (2001).
18. Rosenstiel, T. N., Fisher, A. J., Fall, R. & Monson, R. K. Differential accumulation of dimethylallyl-diphosphate in leaves and needles of isoprene- and methylbutenol-emitting and non-emitting species. *Plant Physiol.* **129**, 1276–1284 (2002).
19. Lichtenthaler, H. K. The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 47–65 (1999).
20. Karl, T. *et al.* On line analysis of the ¹³CO₂ labeling of leaf isoprene suggests multiple subcellular origins of isoprene precursors. *Planta* **215**, 894–905 (2002).
21. Chollet, R., Vidal, J. & O'Leary, M. H. Phosphoenolpyruvate carboxylase: a ubiquitous highly regulated enzyme in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 273–298 (1996).
22. Griffin, K. L. *et al.* Plant growth in elevated CO₂ alters mitochondrial number and chloroplast fine structure. *Proc. Natl Acad. Sci. USA* **98**, 2473–2478 (2001).
23. Wang, X., Lewis, J. D., Tissue, D. T., Seeman, J. R. & Griffin, K. L. Effects of elevated atmospheric CO₂ concentration of leaf dark respiration of *Xanthium strumarium* in light and in darkness. *Proc. Natl Acad. Sci. USA* **98**, 2479–2484 (2001).
24. Crutzen, P. J., Lawrence, M. G. & Poschl, U. On the background photochemistry of tropospheric ozone. *Tellus A* **51**, 123–146 (1999).
25. Sharkey, T. D. & Yeh, S. Isoprene emission from plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 407–436 (2001).
26. Griffin, K. L. *et al.* Leaf respiration is differentially affected by leaf vs. stand-level night-time warming. *Glob. Change Biol.* **8**, 479–485 (2002).
27. Sanadze, G. A., Mgaloblishvili, M. P. & Dalakishvili, K. G. Action of cerulenin, organophosphorous compounds, and synthetic analogs of carbohydrates on CO₂ assimilation and the isoprene effect in isolated Poplar protoplasts. *Sov. Plant Physiol.* **37**, 1–5 (1990).
28. Edwards, G. E., Robinson, S. P., Tyler, N. J. C. & Walker, D. A. A requirement for chelation in obtaining functional chloroplasts of Sunflower and Wheat. *Arch. Biochem. Biophys.* **190**, 421–433 (1978).
29. Magin, N. C., Cooley, B. A., Reiskind, J. B. & Bowes, G. Regulation and localization of key enzymes during the induction of Kranz-less, C4-type photosynthesis in *Hydrilla verticillata*. *Plant Physiol.* **115**, 1681–1689 (1997).
30. Jenkins, C. L. D., Furbank, R. T. & Hatch, M. D. Inorganic carbon diffusion between C4 mesophyll and bundle sheath cells. *Plant Physiol.* **91**, 1356–1363 (1989).

Acknowledgements We thank R. Murthy for assistance at the Biosphere 2 Center, W. Kahtri for technical assistance, and C. Jenkins for supplying DCDP. This work was supported in part by grants from the National Science Foundation (NSF), the Department of Energy and the Packard Foundations to R.F., R.K.M. and K.L.G., an NSF fellowship to M.J.P., a Cooperative Institute for Research in Environmental Sciences fellowship to T.N.R., and a Columbia University Biosphere 2 grant to R.K.M.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.N.R. (e-mail: todd.rosenstiel@colorado.edu).

Speciation along environmental gradients

Michael Doebeli* & Ulf Dieckmann†

* Departments of Zoology and Mathematics, University of British Columbia, Vancouver, Canada V6T 1Z4

† Adaptive Dynamics Network, International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

Traditional discussions of speciation are based on geographical patterns of species ranges^{1,2}. In allopatric speciation, long-term geographical isolation generates reproductively isolated and spatially segregated descendant species^{1,3}. In the absence of geographical barriers, diversification is hindered by gene flow^{1,3,4}. Yet a growing body of phylogenetic and experimental data suggests that closely related species often occur in sympatry

or have adjacent ranges in regions over which environmental changes are gradual and do not prevent gene flow^{5–14}. Theory has identified a variety of evolutionary processes that can result in speciation under sympatric conditions^{15–25}, with some recent advances concentrating on the phenomenon of evolutionary branching^{18,23–25}. Here we establish a link between geographical patterns and ecological processes of speciation by studying evolutionary branching in spatially structured populations. We show that along an environmental gradient, evolutionary branching can occur much more easily than in non-spatial models. This facilitation is most pronounced for gradients of intermediate slope. Moreover, spatial evolutionary branching readily generates patterns of spatial segregation and abutment between the emerging species. Our results highlight the importance of local processes of adaptive divergence for geographical patterns of speciation, and caution against pitfalls of inferring past speciation processes from present biogeographical patterns.

We extended generic Lotka–Volterra models for frequency-dependent competition to individual-based stochastic models of populations occupying a continuous spatial area. In these models, individuals vary in a quantitative trait u , for example a morphological, behavioural or physiological character. In addition, each individual is characterized by its spatial location (x, y) in a square spatial area. In this area resources are distributed such that for each spatial location (x, y) there is a phenotype u_0 with maximal carrying capacity. This phenotype varies linearly with one spatial dimension, $u_0(x) = ax + b$, which represents the most gradual environmental structure possible; the other spatial dimension y is ecologically neutral (Fig. 1). Such a resource gradient in one spatial dimension could, for example, represent variation in temperature, humidity, soil nutrients, or prey items along an altitudinal gradient. Accordingly, the carrying capacity K depends on the phenotype as well as on the spatial location, and is assumed to be of gaussian form: if $N_\sigma(x) = \exp(-\frac{1}{2}x^2/\sigma^2)$ denotes a gaussian function with variance σ^2 , then $K = K_0 \cdot N_{\sigma_K}(u - u_0(x))$. At each spatial location, carrying capacity decreases with phenotypic distance from its maximum at u_0 ; the width of this peak is given by σ_K .

We assumed that the strength of competition between two individuals depends on their phenotypic distance, so that competition is strongest between individuals with similar phenotypes, as, for example, when similarly sized individuals prefer similar types of food. We also assumed that the strength of competition decreases with spatial distance between individuals. Thus, in our individual-based models, the effective population size determining the death rate of a given individual owing to competition depends both on the number of other individuals in its neighbourhood and on their phenotypes. Specifically, in our models, the relative strength of competition between two individuals with phenotypic distance Δu

and spatial distance d is proportional to a product of gaussian functions: $N_{\sigma_c}(\Delta u) \cdot N_{\sigma_s}(d)$. The parameters σ_c and σ_s determine how fast the intensity of competition declines with phenotypic and spatial distance, respectively. In particular, small values of σ_s ensure that severe competition is only felt from individuals that are close-by in the spatial arena.

Finally, to describe spatial movement we assumed that individuals move around in the spatial arena with mean movement distances σ_m and at rates that are independent of location or phenotype. Movement over short distances and localized ecological interactions between individuals allow the population to become spatially structured, whereas frequent movement over long distances tends to result in well-mixed and hence spatially unstructured populations.

Based on these ecological determinants, the evolutionary dynamics of the quantitative trait u were investigated first in asexual populations by allowing for small mutations during birth events. For various reasons (see Methods), no reliable analytical or even deterministic theory for the resultant evolutionary dynamics is available at present, so that the direct investigation of individual-based spatially explicit models is necessary. Evolutionary branching of traits determining competitive interactions has already been studied extensively in analytically tractable models without spatial structure^{18,24,25}, which can be recovered from our models by setting all spatial coordinates to 0. In particular, in the non-spatial version of the adaptive dynamics of the quantitative trait u , evolutionary branching occurs if the strength of competition decreases faster than the resource abundance with phenotypic distance from the capacity-maximizing phenotype, that is, if $\sigma_c < \sigma_K$ (ref. 18).

Critical aspects of spatial structure are determined by the steepness of the environmental gradient and the movement distance. If the gradient is shallow, the environment becomes essentially spatially homogenous. If movement distances are large, the population becomes well-mixed and hence spatially unstructured. In either of these cases the system's behaviour approaches that of the non-spatial model. In particular, evolutionary branching then occurs for the same parameter combinations as in the non-spatial model, that is, for $\sigma_c < \sigma_K$. When evolutionary branching does occur under such conditions, the two evolving phenotypic clusters are scattered randomly over space.

The system's behaviour is dramatically different if the environmental gradient is steep enough and movement distances are short. Evolutionary branching is then accompanied by spatial segregation of the diverging phenotypic clusters (Fig. 2a). Thus, in spatially structured populations evolutionary branching driven by localized and frequency-dependent ecological interactions can generate a sharp phenotypic abutment across a linear environmental gradient. Depending on parameter values, it is possible to observe more than

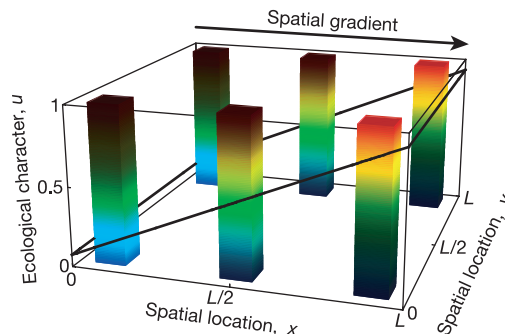


Figure 1 Environmental gradient in carrying capacities. Bright colours correspond to phenotypes that maximize local carrying capacity; these gradually change with spatial location in the x -direction, while the y -direction is ecologically neutral. At any given

location, the carrying capacity decreases with phenotypic distance from the capacity-maximizing phenotype (indicated by diminished brightness).

two distinct and spatially segregated lineages (not shown).

A second and perhaps more important effect is that with significant environmental gradients and short movement distances, evolutionary branching occurs for a much wider range of parameters than in the non-spatial models, that is, for values of σ_c that are much larger than σ_K . The degree to which spatial structure facilitates branching, and the abrupt onset of this facilitation as a function of movement distances, are surprising (Fig. 3a). If movement distances exceed a certain threshold value, parameter requirements for branching in the spatial and non-spatial models are almost exactly the same. However, as movement distances are decreased below this threshold, parameter requirements in the spatial model are suddenly and drastically less restrictive than in the non-spatial model. In fact, if the width σ_s of the spatial interaction kernel is sufficiently small, as is the case in Fig. 3a,

evolutionary branching occurs independently of σ_c .

The mechanisms generating these effects can be understood as follows. An environmental gradient initially induces gradual spatial differentiation due to local adaptation along the gradient. Thus, local adaptation results in a correlation between spatial location and phenotype. When, as assumed here, significant competition only occurs between individuals that are spatially sufficiently close, this correlation decreases the strength of competition between phenotypically distant individuals, and hence increases the degree of frequency dependence in the system. This effect tends to disappear if local adaptation is very incomplete owing to gene flow along shallow gradients, or if dissimilar phenotypes are spatially close because of local adaptation along a very steep environmental gradient. Therefore, facilitation of evolutionary branching due to gradient-induced frequency dependence is expected to be highest

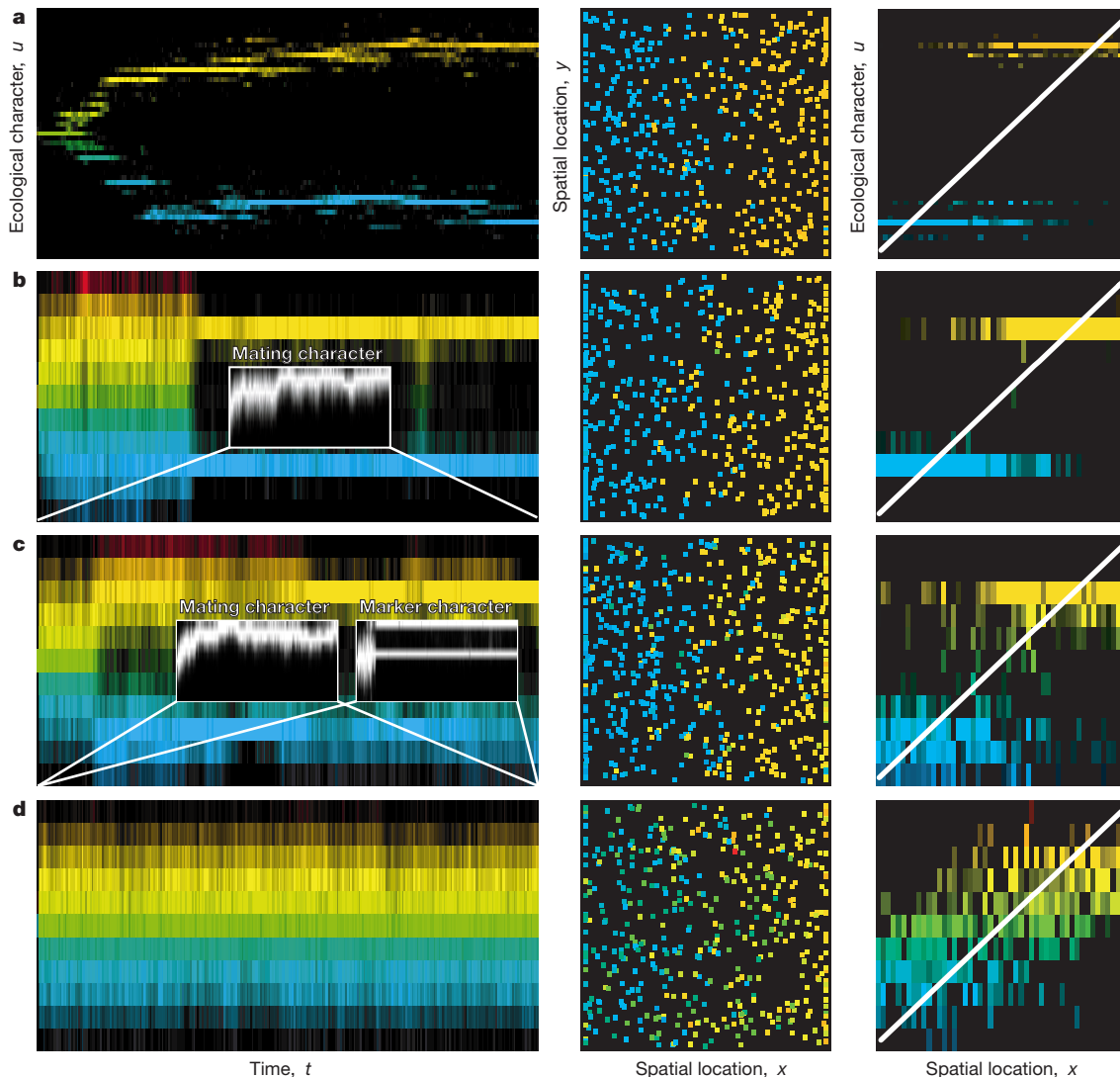


Figure 2 Evolutionary dynamics of adaptive divergence. In each row the left panels show the distribution of phenotypes as a function of time (same colour scheme as in Fig. 1). The middle panels show the distribution of phenotypes across space at the end of the time series, and the right panels show the corresponding frequency distribution of phenotypes as a function of spatial x -location. The white lines indicate the environmental gradient (Fig. 1). **a**, Evolutionary branching with spatial segregation in an asexual population. See Methods for parameter values. Note that $\sigma_c = 2.5\sigma_K$, hence no branching would be expected in the corresponding non-spatial model. **b**, Evolutionary branching with spatial segregation in a sexual population with the same parameter values as in **a** and with assortative mating based on ecological similarity. The evolution of the degree of

assortativeness is shown as an insert in the left panel (intermediate values of the mating character correspond to random mating, low values to disassortative mating, high values to assortative mating). **c**, Evolutionary branching with spatial segregation in a sexual population with assortative mating based on a marker trait. The evolution of assortativeness, as well as the branching in the marker trait, are shown as inserts in the left panel (the two marker branches are in linkage disequilibrium with the two branches of the ecological trait). Parameter values are as given in Methods, except for $\sigma_c = 0.5$, $\sigma_s = 0.3$ and $\tilde{\sigma}_m = 0.38$. **d**, Evolution of a phenotypic gradient in a sexual population with random mating. Parameter values are the same as in **b**, except that random mating with respect to phenotype was enforced.

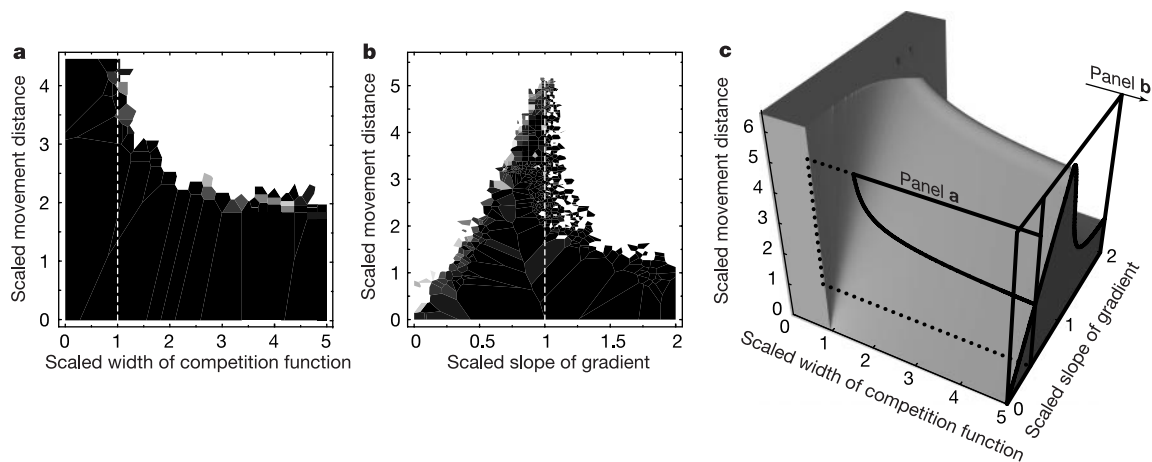


Figure 3 Requirements for spatial evolutionary branching in asexual populations. The model's three dimensionless parameters (see Methods) are displayed on all axes: scaled width of competition function σ_c/σ_K , scaled movement distance $\tilde{\sigma}_m/\sigma_s$, and scaled slope of gradient $a\sigma_s/\sigma_K$. The first two panels show a subdivision of parameter space into polygons (Voronoi tessellation based on simulation data), shaded according to the recorded time to evolutionary branching: black corresponds to branching within the first 500 generations, white corresponds to no branching after 5,000 generations, and shades of grey correspond to branching between generations 500 and 5,000 (including multiple branching, which occurs for very small movement distances). **a**, Effect of direct frequency

dependence. Variation of time until branching with scaled width of competition function and scaled movement distance for asexual populations (at $a\sigma_s/\sigma_K = 0.425$). In non-spatial models^{18,24} only conditions to the left of the dashed line would be expected to induce branching. **b**, Effect of gradient-induced frequency dependence. Variation of time until branching with scaled slope of gradient and scaled movement distance for asexual populations (at $\sigma_c \gg \sigma_K$). In non-spatial models no branching would be expected at all. **c**, Complete characterization of the asexual model, obtained by extrapolating from **a** and **b** and from additional numerical simulations. Evolutionary branching occurs for parameters within the shaded block. The positions of slices in panels **a** and **b** are indicated.

for intermediate environmental gradients. This is demonstrated in Fig. 3b for a case in which the width σ_c of the phenotypic interaction is very large, so that frequency dependence is entirely due to localized interactions between spatially differentiated individuals, and no evolutionary branching at all is expected in the non-spatial model. Figure 3c shows the full characterization of the branching behaviour in asexual populations as a function of the three essential parameters of the model (see Methods). It illustrates that for a range of intermediate environmental gradients, evolutionary diversification is greatly facilitated once movement distances or rates fall below a critical level.

We have extended the spatially structured models to sexual populations in which the quantitative character u is determined additively by several diploid loci (see Methods). The ecological processes remain the same, but instead of reproducing asexually individuals now choose partners within a given spatial neighbourhood (see Methods). If mating is random with respect to phenotypes, evolutionary branching does not occur anymore, regardless of the choice of parameters. Just as in the non-spatial models, random mating brings about recombination between extremal phenotypes, which prevents the evolution of phenotypic bimodality¹⁸. By contrast, evolutionary branching is possible in spatially structured sexual populations if one allows for the evolution of assortative mating. Assortative mating can be based on similarity in the primary character determining the ecological interactions, or it can be based on a selectively neutral marker trait¹⁸. In the latter case, a linkage disequilibrium between the marker trait and the primary trait must evolve for evolutionary branching to occur¹⁸.

Results for spatially structured sexual populations with assortative mating are in general agreement with those from the asexual models. When evolutionary branching occurs in sexual populations, the emerging phenotypic clusters are essentially reproductively isolated because mating is assortative, and the spatial gradient again separates the emerging species into a pattern of spatial segregation (Fig. 2b, c). Such geographical differentiation in the presence of a spatial gradient has been previously observed in a model for competition between two species²⁶; this model did not, however, address the question of speciation. It is important to note

that, if divergence is prevented by random mating, one expects the evolution of a phenotypic gradient along the environmental gradient^{2,27,28}. Such a phenotypic gradient does indeed evolve in our sexual models when individuals mate randomly with regard to their phenotypes (Fig. 2d). Thus, the speciation processes described here are ultimately due to evolutionary branching and to the evolution of assortative mating caused by frequency-dependent interactions under conditions of ecological contact. In sexual populations, speciation through spatial evolutionary branching again occurs for a much wider range of parameters than in the corresponding non-spatial sexual populations, and is most likely with environmental gradients of intermediate slope. The behaviour of the sexual model in which assortative mating is based on similarity in the ecological trait is summarized in Fig. 4. Notice that parameter requirements for spatial evolutionary branching in sexual populations are generally stricter than for asexual populations.

Our results show that intrinsically sympatric processes of adaptive speciation can generate sharp geographical differentiation in the absence of abrupt environmental changes^{2,21}. In traditional models of parapatric speciation due to isolation by distance^{15,22}, diversification is driven by divergent local adaptation or genetic drift in spatially distant locations and is hindered by gene flow. In contrast, our models show that ecological contact may in fact be the driving force for parapatric speciation. Gene flow is of course still a hindrance to local divergence, but the mechanisms generating local disruptive selection require ecological contact. Local disruptiveness in turn selects for assortative mating, which reduces and eventually eliminates gene flow between the emerging species. The latter process is akin to reinforcement²², except that selection for prezygotic isolation emerges dynamically from frequency-dependent ecological interactions and is not a consequence of secondary contact. Spatial segregation between closely related species^{7,13,29} is often used to infer allopatric speciation processes. However, our results show that this inference of process from pattern may be misleading, and that instead the origin and distribution pattern of species abutments is consistent with spatial evolutionary branching. This perspective may be important for understanding the origin of species abutments such as those reported for giant senecios along

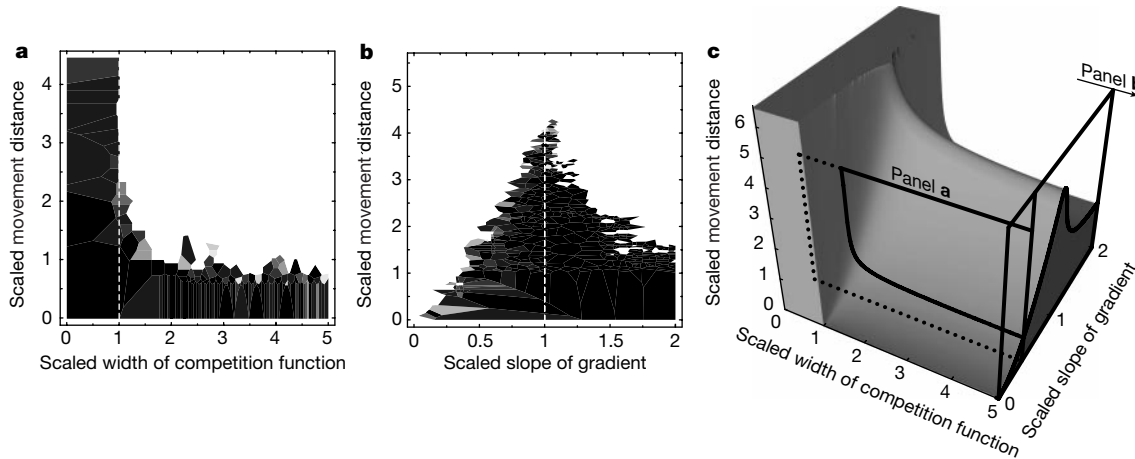


Figure 4 Requirements for spatial evolutionary branching in sexual populations. Same as Fig. 3, but for sexual populations in which assortative mating based on ecological similarity can evolve.

altitudinal gradients⁸, and the origin of hybrid zones such as those between intertidal snails^{7,13}.

Our results also show that gradual spatial structure, potentially even more so than complete spatial isolation, facilitates speciation, because local adaptation along an environmental gradient increases the degree of frequency dependence in spatially localized ecological interactions, and hence the likelihood that these interactions generate disruptive selection. Interestingly, this facilitation is most pronounced for environmental gradients of intermediate slope, a result that is fundamentally different from those expected in classical parapatric speciation models. Other things being equal, we can thus hypothesize that speciation rates are highest in populations exposed to environmental gradients with slopes around σ_K/σ_s (Figs 3b, c and 4b, c), a measure that varies widely between taxa. This conjecture appears to be testable empirically, both by comparative analyses and by studying the effects of environmental gradients on the experimental evolution of diversity in microorganisms¹⁰. In summary, the theory presented here offers a new perspective on the importance of geographical structure for the evolution of diversity by showing that spatially localized interactions along environmental gradients can facilitate speciation through frequency-dependent selection and result in patterns of geographical segregation between the emerging species. □

Methods

Events

Individuals are assigned birth rates b_i , death rates d_i , and movement rates m_i , $i = 1, \dots, n$, where n is the current population size; these rates are updated as necessary after each event. Time is continuous and generations are overlapping. Based on the total rates $B = \sum_{i=1}^n b_i$, $D = \sum_{i=1}^n d_i$, $M = \sum_{i=1}^n m_i$, and $E = B + D + M$, the time lapse until the next event is drawn from an exponential probability distribution with mean $1/E$; the type of event is chosen according to the probabilities B/E , D/E , and M/E . The affected individual i is then chosen with probability b_i/B , d_i/D , or m_i/M , and the chosen individual either gives birth to one offspring, dies, or moves, depending on the event type occurring.

Phenotypes

In the asexual models, ecological phenotypes $0 \leq u \leq 1$ vary continuously. In the sexual models, ecological, mating and marker phenotypes are each determined by l diallelic diploid loci with additive effects and free recombination. Ecological and marker phenotypes vary between 0 and 1. The mating character varies between -1 (disassortative mating) and $+1$ (assortative mating); 0 encodes random mating^{18,24}.

Gradient

Individuals have a spatial location (x, y) , with $0 \leq x, y \leq L$. The carrying capacity for the ecological phenotype u at spatial location (x, y) is $K(u, x, y) = K_0 \cdot N_{\sigma_K}(u - u_0(x))$, where $u_0(x) = a(x - L/2) + L/2$ is the phenotype maximizing K at location x , and $0 \leq a \leq 1$ is the slope of the environmental gradient (Fig. 1); u_0 thus varies over space in the range $[(1 - a)L/2, (1 + a)L/2]$.

Death

The effective density experienced by an individual i with phenotype u at location (x, y) is a weighted sum, $n_{\text{eff}}(u, x, y) = \frac{1}{2\pi\sigma_s^2} \sum_{\Delta u, d} N_{\sigma_s}(\Delta u) \cdot N_{\sigma_d}(d)$ extending over all pairs $(\Delta u, d)$ of phenotypic and spatial distances between the focal and other individuals. The resultant logistic death rate is $d_i = n_{\text{eff}}(u, x, y)/K(u, x, y)$.

Birth

In asexual populations, individuals reproduce at a fixed rate $b_i = b$. Offspring express the parental phenotype unless a mutation occurs at probability μ_a , in which case their phenotype u' is chosen according to $N_{\sigma_s}(u' - u)$. For sexual populations, refs. 18, 24 describe how an individual i slated for reproduction chooses a partner j according to phenotype-based mating probabilities p_{ij} depending on its mating character and the partner's phenotypic distance in either ecological or marker character. Although this does not yet imply a cost to choosiness (the p_{ij} are normalized), there is already a cost to rarity when mating is assortative (individuals with uncommon phenotypes will rarely be chosen as mating partners). In our spatial models, the location-based component q_{ij} of mating probabilities decreases according to $(2\pi\sigma_p^2)^{-1} \cdot N_{\sigma_p}(d)$ with the spatial distance d between potential partners. This induces a cost to preferring locally rare phenotypes: $b_i = b/(1 + c/n_p)$, where $n_p = \sum_{j=1, j \neq i}^n p_{ij} \cdot q_{ij}$ is the number of suitable mating partners locally available to individual i , and c determines the cost's strength. Notice that assortativeness often evolves despite this cost. (For sexual populations, only females are modelled; males are assumed to have the same density and frequency distributions as females. Given the probabilistic recipe for finding mates, this simplification is uncritical and seems justified bearing in mind otherwise even more computationally demanding sex-structured simulations.) After recombination, the offspring genotype is subjected to allelic mutations according to a reversal probability μ_r . Offspring undergo an initial movement event from the location of their parent.

Movement

Individuals move at a fixed rate $m_i = m$, undergoing displacements d in the x - and y -directions drawn independently according to $N_{\sigma_m}(d)$, resulting in an average movement distance σ_m . Boundaries are reflective in the x -direction and periodic in the ecologically neutral y -direction. For aiding interpretation it is convenient to consider the expected movement distance during the average lifespan of an individual at demographic equilibrium, $\tilde{\sigma}_m = \sqrt{(m/b) \cdot \sigma_m}$ (where b is the birth rate).

Parameters

Unless otherwise stated: $l = 10$, $L = 1$, $K_0 = 500$, $\sigma_K = 0.3$, $a = 0.95$, $\sigma_c = 0.9$, $\sigma_s = 0.19$, $b = 1$, $\mu_a = 0.005$, $\sigma_a = 0.05$, $\sigma_p = 0.2$, $c = 10$, $\mu_r = 0.001$, $m = 5$, $\tilde{\sigma}_m = 0.27$. In the salient limit of large L , K_0 and small $\mu_a \sigma_a^2$ the asexual model has no more than three essential dimensionless parameters: σ_c/σ_K , $\tilde{\sigma}_m/\sigma_s$, and $a\sigma_s/\sigma_K$ (obtained from choosing units for time, space, and phenotype as $1/b$, σ_s , and σ_K). This important simplification allows for a complete characterization of the asexual model as shown in Fig. 3a–c. To illustrate the biological meaning of the three dimensionless parameters, we note that if the first, σ_c/σ_K , is equal to 1, then the phenotypic distance reducing the strength of competition by a given amount is the same as the phenotypic distance from the capacity-maximizing phenotype reducing the carrying capacity by the same amount; if the second, $\tilde{\sigma}_m/\sigma_s$, is equal to 1, then the expected movement distance during an average lifespan equals σ_s , the width of the spatial interaction kernel; and if the third, $a\sigma_s/\sigma_K$, is equal to 1, then movement of a capacity-maximizing phenotype by σ_s in the x -direction reduces its carrying capacity by $1/e$.

Approximations

We also investigated approximations of the asexual individual-based model. Although

being more tractable by deterministically describing the dynamics of a population density $n(x,u)$, these approximations are problematical. First, such continuum approximations are based on the limit of infinite local population sizes (local both in x and u), which is even more difficult to justify biologically than the limit of infinite global population size, widely used in population ecology. Second, a conveniently simple reaction-diffusion approximation of this system, derived for small σ_m and σ_s , is dynamically unstable. Third, these approximations ignore the implications of reproductive (and other) pair correlations and local density fluctuations, both of which have been shown to critically affect ecological and evolutionary dynamics³⁰. Fourth, the deterministic approximation blurs the sharp bifurcation boundary in Fig. 3a and is also inaccurate in predicting the boundary's location. Fifth, extending the deterministic approximation to multilocus genetics is not feasible without incurring further unjustified assumptions.

Received 12 July; accepted 18 October 2002; doi:10.1038/nature01274.

- Mayr, E. *Animal Species and Evolution* (Harvard Univ. Press, Cambridge, Massachusetts, 1963).
- Endler, J. A. *Geographic Variation, Speciation, and Clines* (Princeton Univ. Press, Princeton, New Jersey, 1977).
- Coyne, J. A. Genetics and speciation. *Nature* **355**, 511–515 (1992).
- Felsenstein, J. Scepticism towards Santa Rosalia, or why are there so few kinds of animals. *Evolution* **35**, 124–138 (1981).
- Schliwien, U. K., Tautz, D. & Pääbo, S. Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* **368**, 629–632 (1994).
- Schluter, D. Experimental evidence that competition promotes divergence in adaptive radiation. *Science* **266**, 798–801 (1994).
- Johannesson, K., Rolán-Alvarez, E. & Ekendahl, A. Incipient reproductive isolation between two sympatric morphs of the intertidal snail *Littorina saxatilis*. *Evolution* **49**, 1180–1190 (1995).
- Knox, E. B. & Palmer, J. D. Chloroplast DNA variation and the recent radiation of giant senecios (Asteraceae) on the tall mountains of Eastern Africa. *Proc. Natl Acad. Sci. USA* **92**, 10349–10353 (1995).
- Feder, J. L. *Endless Forms* (eds Howard, D. J. & Berlocher, S. H.) 130–144 (Oxford Univ. Press, Oxford, UK, 1998).
- Rainey, P. B. & Travisano, M. Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69–72 (1998).
- Gislason, D., Ferguson, M., Skúlason, S. & Snorrason, S. S. Rapid and coupled phenotypic and genetic divergence in Icelandic Arctic char (*Salvelinus alpinus*). *Can. J. Fish. Aquat. Sci.* **56**, 2229–2234 (1999).
- Schliwien, U. K., Rassmann, K., Markmann, M., Markert, J. & Tautz, D. Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. *Mol. Ecol.* **10**, 1471–1488 (2001).
- Wilding, C. S., Butlin, R. K. & Grahame, J. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *J. Evol. Biol.* **14**, 611–618 (2001).
- Via, S. Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol. Evol.* **16**, 381–390 (2001).
- Lande, R. Rapid origin of sexual isolation and character divergence in a cline. *Evolution* **36**, 213–223 (1982).
- Kawecki, T. J. Sympatric speciation via habitat specialization driven by deleterious mutations. *Evolution* **51**, 1751–1763 (1997).
- Kondrashov, A. S. & Kondrashov, F. A. Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**, 351–354 (1999).
- Dieckmann, U. & Doebeli, M. On the origin of species by sympatric speciation. *Nature* **400**, 354–357 (1999).
- Van Doorn, G. S., Luttikhuisen, P. C. & Weissing, F. J. Sexual selection at the protein level drives the extraordinary divergence of sex-related genes during sympatric speciation. *Proc. R. Soc. Lond. B* **268**, 2155–2161 (2001).
- Gavrilets, S. & Waxman, D. Sympatric speciation by sexual conflict. *Proc. Natl Acad. Sci. USA* **99**, 10533–10538 (2002).
- Mizera, F. & Meszén, G. Spatial niche packing, character displacement and adaptive speciation in an environmental gradient. *Evol. Ecol. Res.* (in press).
- Turelli, M., Barton, N. H. & Coyne, J. A. Theory and speciation. *Trends Ecol. Evol.* **16**, 330–343 (2001).
- Geritz, S. A. H., Kisdi, E., Meszén, G. & Metz, J. A. J. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* **12**, 35–57 (1998).
- Doebeli, M. & Dieckmann, U. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *Am. Nat.* **156**, S77–S101 (2000).
- Kisdi, E. Evolutionary branching under asymmetric competition. *J. Theor. Biol.* **197**, 149–162 (1999).
- Case, T. J. & Taper, M. L. Interspecific competition, environmental gradients, gene flow, and the coevolution of species' borders. *Am. Nat.* **155**, 583–605 (2000).
- Slatkin, M. Spatial patterns in the distribution of polygenic characters. *J. Theor. Biol.* **70**, 213–228 (1978).
- Barton, N. H. Clines in polygenic traits. *Genet. Res.* **74**, 223–236 (1999).
- Barracough, T. G. & Vogler, A. P. Detecting the geographical pattern of speciation from species-level phylogenies. *Am. Nat.* **155**, 419–434 (2000).
- Dieckmann, U., Law, R. & Metz, J. A. J. *The Geometry of Ecological Interactions: Simplifying Spatial Complexity* (Cambridge Univ. Press, Cambridge, UK, 2000).

Acknowledgements We thank H. Metz, D. Tautz, G. Meszén, D. Schluter, E. Knox, O. Leimar, M. Kirkpatrick and T. Barraclough for discussions and comments. The order of authors is reverse alphabetical.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.D. (e-mail: doebeli@zoology.ubc.ca).

Loss and recovery of wings in stick insects

Michael F. Whiting*, Sven Bradler† & Taylor Maxwell‡

* Department of Integrative Biology, Brigham Young University, Provo, Utah 84602, USA

† Institut für Zoologie und Anthropologie, Georg August Universität Göttingen, Germany

‡ Washington University, Department of Biology, St Louis, Missouri 63130, USA

The evolution of wings was the central adaptation allowing insects to escape predators, exploit scattered resources, and disperse into new niches, resulting in radiations into vast numbers of species¹. Despite the presumed evolutionary advantages associated with full-sized wings (macroptery), nearly all pterygote (winged) orders have many partially winged (brachypterous) or wingless (apterous) lineages, and some entire orders are secondarily wingless (for example, fleas, lice, grylloblattids and mantophasmatids), with about 5% of extant pterygote species being flightless^{2,3}. Thousands of independent transitions from a winged form to winglessness have occurred during the course of insect evolution; however, an evolutionary reversal from a flightless to a volant form has never been demonstrated clearly for any pterygote lineage. Such a reversal is considered highly unlikely because complex interactions between nerves, muscles, sclerites and wing foils are required to accommodate flight⁴. Here we show that stick insects (order Phasmatodea) diversified as wingless insects and that wings were derived secondarily, perhaps on many occasions. These results suggest that wing developmental pathways are conserved in wingless phasmids, and that 're-evolution' of wings has had an unrecognized role in insect diversification.

Stick insects are large terrestrial insects that exhibit extreme forms

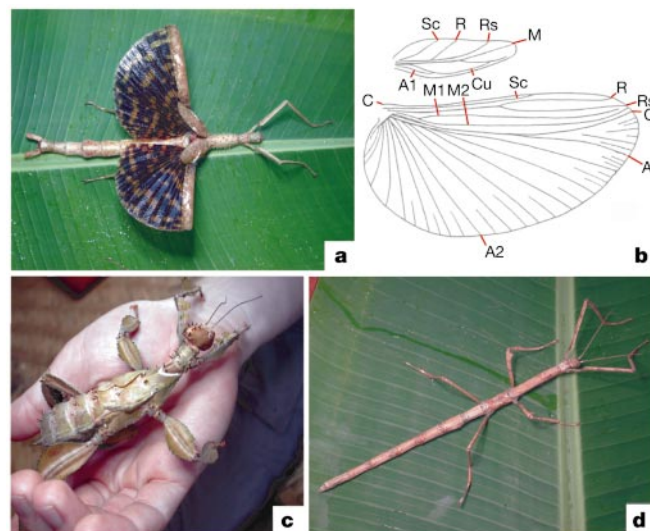


Figure 1 Examples of wing features in stick insects. **a**, Example of a fully winged (macropterous) female phasmid (*Phasma gigas*) with enlarged hindwings and thickened forewings. **b**, Wing venation of male *Phyllium celebicum* with major veins labelled, demonstrating homology with other insect wing veins. A, anal vein; C, costa vein; Cu, cubitus vein; M, medial vein; R, radius vein; Rs, radial sector vein; Sc, subcosta vein. **c**, Example of a partially winged (brachypterous) female phasmid (*Extatosoma popa*) with reduced hindwings. **d**, Example of a wingless (apterous) female phasmid (*Leprocailinus* sp.) with wings entirely absent.

of morphological and behavioural crypsis as mimics of sticks and leaves. Phasmids are chiefly arboreal insects, but a few of the more robust taxa (for example, *Eurycantha*, *Agathemera*, Heteropterygidae) occur primarily near the ground, in the leaf litter. The monophyly of Phasmatodea is supported by a series of distinctive morphological characters including prothoracic repellent glands, absence of mitochondria in spermatozoa, and male vomer^{5,6}. Phasmids belong among the basal winged insect orders (Polyneoptera), but their sister group is unknown, although Caelifera (grasshoppers)⁷, Orthoptera⁸, Dermaptera (earwigs)⁶, Grylloblattoidea

and Dermaptera⁹, Dictyoptera (mantids, cockroaches and termites)¹⁰, and Embiidina (web spinners)^{6,11} are all candidates. It is difficult to decipher the phylogenetic relationships within Phasmatodea because of the convergent morphology associated with their remarkable crypsis. Current classification is based on some dubious morphological characters¹², and no formal investigation of phasmid phylogeny has yet been published.

Of the 3,000 described species of phasmids placed in 3 families and approximately 500 genera, only 40% are fully winged, with the remainder being partially winged or wingless. In fully winged

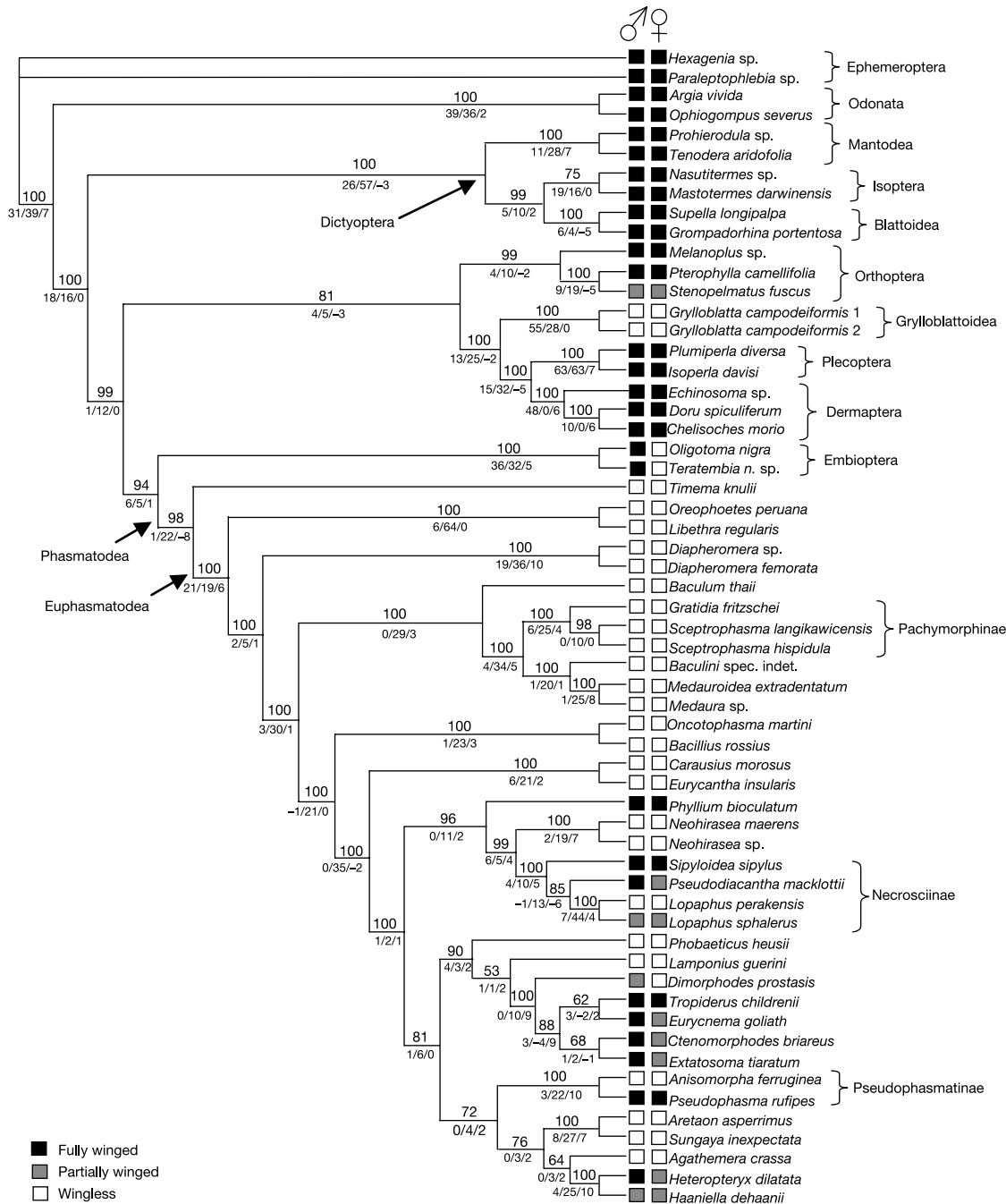


Figure 2 Phylogeny of Phasmatodea on the basis of molecular data. Shown is the single optimization alignment tree based on 18S rDNA, 28S rDNA and histone 3. Nonparametric bootstrap supports are given above each node and partitioned Bremer supports are below each node in the order 18S/28S/H3. This topology is congruent with the maximum

likelihood and bayesian topology (see Supplementary Information). Boxes at the end of nodes represent wing character states for males and females, respectively. n. sp., new species; spec. indet., undetermined species.

phasmids, the front wing is reduced and thickened, and the hind wing is enlarged with an extensive anal region (Fig. 1a). Vein homology is readily assessed between phasmids and other insect groups, and analysis indicates that phasmid wings are homologous to those of other insects (Fig. 1b). Partially winged phasmids have reduced hindwings with a small anal region, and individuals are not capable of sustained flight (Fig. 1c). In wingless phasmids, no wing remnants are present (Fig. 1d). The extent of wing development of male phasmids is never less than females: if the male is wingless, the female is always wingless; if the male is partially winged, the female is partially winged or wingless, but never fully winged; if the male is fully winged, the female may be fully winged, partially winged or wingless.

DNA sequence data were used to place the Phasmatodea among the polyneopterous insect orders and to estimate phylogenetic relationships for major stick insect lineages. The entire regions of 18S ribosomal DNA (18S rDNA; about 1,900 base pairs (bp)), 28S rDNA (2,250 bp), and a portion of histone 3 (H3, 372 bp) were sequenced from 22 outgroup and 37 ingroup taxa representing all Polyneoptera and 14 of the 19 recognized phasmid subfamilies. Trees were reconstructed through optimization alignment and nodal support determined by nonparametric bootstrap and partitioned Bremer support (see Methods and Fig. 2). Partitioned Bremer supports indicate that 31% of the signal is provided by 18S, 62% by 28S and 7% by H3. Overall, H3 provides resolution at the shallow nodes, 18S at the deep nodes, and 28S throughout the

entire topology. Most of the nodes are very well supported, with 44 out of 56 nodes with bootstrap values of 95 or higher and 52 out of 56 nodes with a total Bremer support of 5 or greater. These data support the monophyly of each insect order, the basal placement of Timematidae, and the monophyly of the groups Euphasmatodea, Pachymorphinae, Necrosiinae, Pseudophasmatinae and Lanceocercata, in agreement with other studies^{5,6,8,13}. Analyses using maximum parsimony, maximum likelihood, and bayesian methods result in a topology that is highly congruent with the optimization alignment topology (see Supplementary Information).

The most parsimonious reconstruction (MPR) of the states 'winged' and 'wingless' unambiguously places the ancestral state for phasmids as wingless, with wings derived on four (ACCTRAN optimization) or five (DELTRAN optimization) occasions (Fig. 3). Forcing the ancestral state as winged requires 13 steps in males and 14 steps in females, and is the MPR only when wing gain is weighted six times wing loss. Similar results are found when winged is divided into 'fully winged' and 'partially winged' and the states are treated as unordered. Wing states were also mapped by means of likelihood methods on both the bayesian and likelihood trees, to take into account branch lengths in determining probabilities of ancestral states¹⁴. When the rate of wing gain is set equal to that of wing loss, the phasmid ancestor is reconstructed as wingless ($P < 0.001$), with four independent wing gains in more derived stick insect lineages. When the ratio of the rate of wing loss over wing gain ranges from 1 to 1,400, there is a >95% probability that the ancestral phasmid was wingless.

These results support the hypothesis that the ancestral condition in Phasmatodea is wingless, that the first six basal phasmid lineages are entirely wingless, and that fully developed wings were derived later in phasmid evolution, on as many as four occasions. Clearly, the presence of wings is a very plastic feature in phasmids, with congeneric species (for example, *Lopaphus*) exhibiting both partially winged and wingless states. One of the correlates with winglessness in insects is increased female fecundity^{2,3}, and as phasmids scatter specially modified eggs individually rather than concentrating them in large numbers similar to their sister taxon, there may have been a selective advantage early in phasmid evolution to shift to winglessness to facilitate fecundity and increased crypsis. The detailed homology in wing features shared among phasmids and other insects suggests that wings did not re-evolve *de novo* in phasmids, but are rather a re-expression of the basic insect wing which was lost in ancestral stick insects. Entomologists have long assumed that re-evolution of wings in apterous lineages was impossible, because functional wings require complex interactions among multiple structures, and the associated genes would be free to accumulate mutations in wingless lineages, effectively blocking the path for any future wing reacquisition. However, this assumption requires that developmental pathways for wing formation are largely independent of pathways required for development of other structures. For instance, in *Drosophila* and other insects, leg and wing imaginal discs have a common origin from a single group of cells and the developmental pathway for wing determination has been largely co-opted (recruited) from the pathway required for limb formation^{15,16}. Therefore it is not surprising that the basic genetic instructions for wing formation are conserved in wingless insects, because similar instructions are required to form legs, and probably other critical structures¹⁶. Studies of flight motor patterns in flying and non-flying phasmids indicate that the non-flying phasmids have retained the neural structures and basic functional circuitry required for flight, as indicated by flight-specific neural activity in thoracic muscles¹⁷, demonstrating that the loss of wings does not correlate with the loss of flight musculature and innervation. Wing development depends on multiple gene systems, transcription factors, secreted proteins, and receptors¹⁵, and mutations in any one of these factors may lead to winglessness. Given the multitude of factors involved in wing

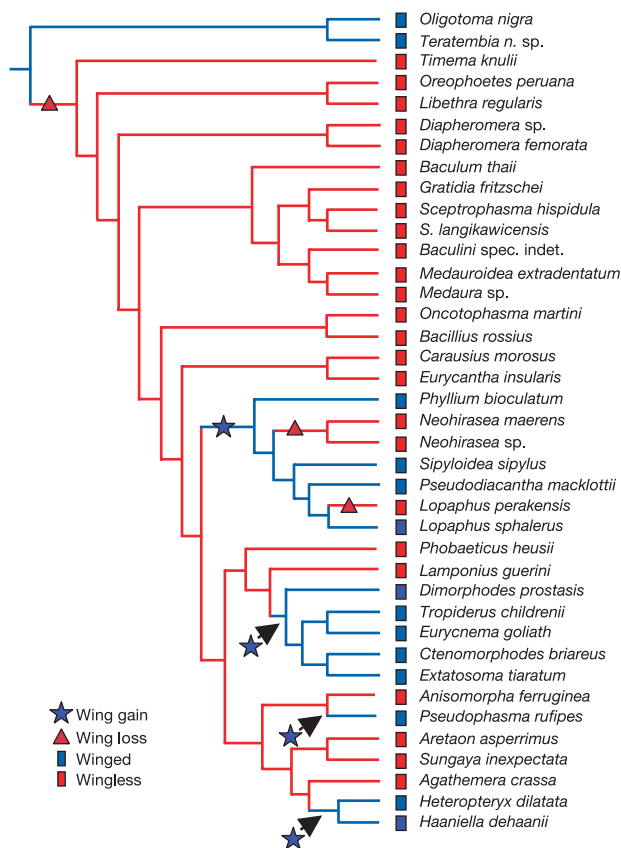


Figure 3 Character mapping of wing types on phasmid phylogeny. Parsimony optimization (ACCTRAN) of winged (blue) and wingless (red) states for male phasmids on the optimization alignment topology. This reconstruction requires seven steps with four wing gains and three losses; DELTRAN optimization requires five wing gains and two losses. Maximum likelihood reconstruction produces similar results (see Supplementary Information).

formation, it seems probable that the specific cause for winglessness will differ from lineage to lineage, but that the basic blueprint for wing formation can remain largely intact, even over large evolutionary time periods.

Re-evolving complex structures from the same basic building blocks may be a more general trend in evolution than previously recognized. For example, it is plausible that all modern animals with eyes evolved from a common ancestor that possessed a primitive image-forming organ¹⁸, controlled by the master gene *Pax-6*. Our results support the hypothesis that the developmental pathway for wing formation evolved only once in insect diversification, but that wings evolved many times by silencing and re-expressing this pathway in different lineages during insect evolution. Thus, wing loss does not seem to be an evolutionary dead end, and the ability to regain a wing over evolutionary time means that lineages have the adaptive advantages of being both winged and wingless. The transition from wingless to flying forms may be a common theme in insect evolution, as it has been suspected within one genus of water striders¹⁹, and may occur in cockroaches and Hemiptera, which also exhibit a wide diversity of wing forms. To our knowledge, this is the first example of a complex feature being lost and later recovered in an evolutionary lineage, and it is possible that the reacquisition of complex features may have an important role in evolutionary diversification. □

Methods

Taxon selection and molecular methods

Outgroup exemplars include multiple representatives of Ephemeroptera, Odonata and all polyneopterous insect orders, except for Zoraptera whose assignment to Polyneoptera is more questionable^{6,8}. Phasmid exemplars include all families and 14 out of 19 subfamilies¹², including Bacteriinae (1 sp.), Eurycanthinae (1 sp.), Diapheromerinae (5 spp.), Lonchodinae (3 spp.), Necrosiinae (4 spp.), Pachymorphinae (3 spp.), Phasmatinae (7 spp.), Tropidoderinae (2 spp.), Xeroderinae (1 sp.), Bacillinae (1 sp.), Heteropteryginae (4 spp.), Phyllinae (1 sp.), Pseudophasmatinae (3 spp.) and Timeminae (1 sp.). The five subfamilies left unsampled are mostly minor groups/species-poor taxa. Genomic DNA was extracted from ethanol-preserved specimens from leg muscle tissue using standard protocols¹¹. Genomic DNA templates and controls were amplified by polymerase chain reaction (PCR) in a Perkin-Elmer 9700 thermocycler using primers modified for insects^{20,21}. We monitored product yield, specificity, and potential contamination by agarose gel electrophoresis. Target products were purified and cycle-sequenced using ABI Prism Big Dye Terminator chemistry. Sequencing reactions were column purified and fractionated with the ABI 377 automated sequencer. DNA was sequenced from complementary strands and correction of chromatography data was facilitated by the program Sequencher 3.1.1.

Alignment and phylogenetic analysis

Alignment of histone 3 was based on conservation of amino acid reading frame. A gross alignment was performed on the ribosomal genes by manually aligning the conserved domains across all taxa in Sequencher 3.1.1. Each conserved domain, and the variable regions between the conserved domains, underwent optimization alignment using the computer program POY, run in parallel mode across an IBM SP2 Supercomputer; this is the first implementation, to our knowledge, of this program on a supercomputer. Ribosomal expansion regions were excluded from outgroups but aligned for ingroups as described elsewhere²². Sixteen parameter sets (gap/change ratio 1–4; transition/transversion ratio 1–4) were applied across all data partitions (18S rDNA, 28S rDNA, H3, and total). The parameter combination with gaps, transitions and transversions weighted equally minimized incongruence among data partitions (incongruence length difference (ILD) metric = 0.01722), and was selected as the best-justified parameter set for optimization alignment²³. Topologies from parameter combination sets adjacent to the optimal combination produced similar phylogenetic results. Optimal parameters produced a single topology (length = 9,057, consistency index = 0.508, retention index = 0.618), and the implied alignment from this topology was used to calculate bootstrap and partitioned Bremer support values. Bootstrap values were calculated in PAUP*4.0b10 (ref. 24) with 10,000 replicates, 5 random additions per replicate, and TBR branch swapping. Partitioned Bremer support values were calculated using TreeRot²⁵. The incongruence length difference test²⁶ indicated insufficient evidence to reject the hypothesis of data set congruence ($P = 0.892$), therefore individual data sets were combined into a total evidence analysis. The states 'fully winged', 'partially winged' and 'wingless' were treated as unordered characters and mapped for males and females by means of parsimony using MacClade 4.0 (ref. 27). For the ingroup, the MPR required ten steps (males) and nine steps (females), the ancestral phasmid state was reconstructed unambiguously as wingless, and wings were derived four times unambiguously. Constraining the ancestral phasmid state to fully winged (by forcing state changes as irreversible) required 28 steps for males and 30 steps for females, and is the MPR when wing gain is 11 times wing loss. Fully winged and partially winged were combined into winged and the MPR required 7 steps (males and females), the ancestral state was

reconstructed as wingless, and wings were unambiguously derived on four occasions. Mapping of characters by means of likelihood was performed using the program Discrete²⁸ on topologies generated by means of standard likelihood and bayesian methods (see Supplementary Information). Wings were mapped for males and females (character 1 and 2) with two states: wings present or absent. Character states were mapped on the maximum likelihood topology and the bayesian topology with the rate of wing loss set to a value of 1, and the rate of wing gain set to 20 individual values, spanning the interval 1 to 0.00001. Character states were also mapped on a set ($n = 30$) of bayesian trees²⁹, and the mean probability that the ancestor to Euphasmatodea was wingless is 0.9996 ± 0.000246 .

Received 29 May; accepted 31 October 2002; doi:10.1038/nature01313.

- Hennig, W. *Insect Phylogeny* (Academic, New York, 1981).
- Roff, D. A. The evolution of flightlessness: is history important? *Evol. Ecol.* **8**, 639–657 (1994).
- Wagner, D. L. & Liebherr, J. K. Flightlessness in insects. *Trends Ecol. Evol.* **7**, 216–220 (1992).
- Dudley, R. *The Biomechanics of Insect Flight* (Princeton Univ. Press, Princeton, New Jersey, 2000).
- Tilgner, E. H., Kiselyova, T. G. & McHugh, J. V. A morphological study of *Timema cristinae* Vickery with implications for the phylogenetics of Phasmida. *Dtsch. Entomol. Z.* **46**, 149–162 (1999).
- Kristensen, N. P. in *The Insects of Australia: A textbook for Students and Research Workers*, 2nd edn (eds Naumann, I. D. et al.) 125–140 (CSIRO, Melbourne Univ. Press, Melbourne, 1991).
- Hennig, W. *Die Stammesgeschichte der Insekten* (Kramer, Frankfurt am Main, Germany, 1969).
- Wheeler, W. C., Whiting, M. F., Wheeler, Q. D. & Carpenter, J. M. The phylogeny of the extant hexapod orders. *Cladistics* **17**, 113–169 (2001).
- Kamp, J. W. Numerical classification of the orthopteroids, with special reference to the Grylloblattodea. *Can. Entomol.* **105**, 1235–1249 (1973).
- Beutel, R. G. & Gorb, S. N. Ultrastructure of attachment specializations of Hexapods (Arthropoda): evolutionary patterns inferred from a revised ordinal phylogeny. *J. Zool Syst. Evol. Res.* **39**, 177–207 (2001).
- Whiting, M. F., Carpenter, J. C., Wheeler, Q. D. & Wheeler, W. C. The Strepsiptera problem: phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. *Syst. Biol.* **46**, 1–68 (1997).
- Günther, K. Über die taxonomische Gliederung und geographische Verbreitung der Insektenordnung der Phasmatodea. *Beitr. Ent.* **3**, 541–563 (1953).
- Bradler, S. The Australian stick insects—a monophyletic group within the Phasmatodea? *Zoology* **104**, 69 (2001).
- Pagel, M. The maximum likelihood approach to reconstructing ancestral characters states of discrete characters on phylogenies. *Syst. Biol.* **48**, 612–622 (1999).
- Kim, J. et al. Integration of positional information and identity by *Drosophila vestigial* gene. *Nature* **382**, 133–138 (1996).
- Cohen, B., Simcox, A. A. & Cohen, S. M. Allocation of the thoracic imaginal primordia in the *Drosophila* embryo. *Development* **117**, 597–608 (1993).
- Kutsch, W. & Kittman, R. Flight motor pattern in flying and non-flying Phasmida. *J. Comp. Physiol.* **168**, 483–490 (1991).
- Halder, G., Callaerts, P. & Gehring, W. J. Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. *Science* **167**, 1788–1792 (1995).
- Anderson, N. M. in *The Origin of Biodiversity in Insects: Phylogenetic Tests of Evolutionary Scenarios* (ed. Grandcolas, P.) 91–108, (Mem. Mus. Natl. Hist. Nat., Paris, 1997).
- Colgan, D. J. et al. Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Aust. J. Zool.* **46**, 419–437 (1998).
- Whiting, M. F. Mecoptera is paraphyletic: multiple genes and phylogeny of Mecoptera and Siphonaptera. *Zoologica Scripta* **31**, 93–104 (2002).
- Whiting, M. F. in *Molecular Systematics and Evolution: Theory and Practice* (eds DeSalle, R., Wheeler, W. C. & Giribet, G.) 69–84 (Birkhäuser, Basel, 2002).
- Wheeler, W. Fixed character states and the optimization of molecular sequence data. *Cladistics* **15**, 379–386 (1999).
- Swofford, D. L. PAUP*: Phylogenetic analysis using parsimony (*and other methods) (version 4.0b10) (Sinauer, Sunderland, Massachusetts, 2000).
- Sorenson, M. D. TreeRot (version 2) (Boston Univ. Press, Massachusetts, 1999).
- Farris, J. S., Källersjö, M., Kluge, A. G. & Bult, C. Testing significance of incongruence. *Cladistics* **10**, 315–320 (1994).
- Maddison, D. R. & Maddison, W. P. MacClade 4: Analysis of phylogeny and character evolution (version 4) (Sinauer, Sunderland, Massachusetts, 2000).
- Pagel, M. *Discrete* (Univ. Reading Press, Reading, UK, 1999).
- Lutoni, F., Pagel, M. & Reeb, V. Major fungal lineages are derived from lichen symbiotic ancestors. *Nature* **411**, 937–940 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank E. Tilgner for providing some specimens; A. Whiting and L. Harmon for assistance in data analysis; and J. Cherry, M. Gruwell, M. Terry, H. Ogden, J. Robertson and K. Jarvis for generating sequence data. Analyses were performed in the Fulton Supercomputer Center, Brigham Young University, and parallel software implementation was performed by M. Clement and Q. Snell. DNA sequences are deposited in GenBank under accession numbers AY121129–AY121186 and AY125216–AY125326.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.F.W. (e-mail: Michael_Whiting@byu.edu).

Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes

Kaveh Ashrafi*†, Francesca Y. Chang*, Jennifer L. Watts‡, Andrew G. Fraser§, Ravi S. Kamath§, Julie Ahringer§ & Gary Ruvkun*†

* Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

† Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, USA

‡ Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164, USA

§ Wellcome/Cancer Research UK Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK

Regulation of body fat storage involves signalling between centres that regulate feeding in the brain and sites of fat storage and use in the body^{1,2}. Here we describe an assay for analysing fat storage and mobilization in living *Caenorhabditis elegans*. By using RNA-mediated interference (RNAi)^{3,4} to disrupt the expression of each of the 16,757 worm genes, we have systematically screened the *C. elegans* genome for genes necessary for normal fat storage. We identify 305 gene inactivations that cause reduced body fat and 112 gene inactivations that cause increased fat storage. Analysis of the fat-reducing gene inactivations in insulin, serotonin and tubby signalling mutants of *C. elegans*, which have increased body fat, identifies a core set of fat regulatory genes as well as pathway-specific fat regulators. Many of the newly identified worm fat regulatory genes have mammalian homologues, some of which are known to function in fat regulation. Other *C. elegans* fat regulatory genes that are conserved across animal phylogeny, but have not previously been implicated in fat storage, may point to ancient and universal features of fat storage regulation, and identify targets for treating obesity and its associated diseases.

We used the vital dye Nile Red (5H-benzo[α]phenoxazine-5-one, 9-diethylamino)⁵ to visualize fat storage droplets in living worms (see ref. 27 and Supplementary Information for details of all Methods). Addition of Nile Red to *Escherichia coli*, the laboratory diet of *C. elegans*, resulted in uptake and incorporation of the dye into lipid droplets in intestinal cells, the principal site of worm fat storage (data not shown). The Nile Red pattern of fat staining matched the staining that resulted from feeding *C. elegans* fatty acids that had been covalently labelled with the fluorescent dye BODIPY, and was similar to the pattern of fat deposits previously observed after staining fixed animals with Sudan Black B (ref. 6 and data not shown). Nile Red staining did not affect growth rate, brood size, feeding or lifespan.

Alterations in the fat content of *C. elegans* mutants known to have modified fat storage were detectable by Nile Red staining. A neuroendocrine system comprising parallel insulin-like and transforming growth factor- β (TGF- β) signalling pathways controls whether *C. elegans* larvae grow to fast metabolizing or fat-storing adults⁷. The increased fat stores in animals with decreased *daf-2*(*e1370*) insulin-like signalling^{6,8} were readily detectable in living animals by Nile Red staining (Fig. 1a). Consistent with epistasis analysis of the *daf* pathway, the increased fat of *daf-2* mutant animals was suppressed by mutations in the forkhead transcription factor *daf-16* (refs 9, 10) or in the *PTEN* phosphatase *daf-18* (ref. 11), but not by a mutation in the TGF- β pathway SMAD-like protein *daf-3* (ref. 7, Fig. 1a and Supplementary Fig. 1).

As in mammals, serotonin and tubby signalling pathways affect *C. elegans* body fat. Mice deficient either in *Tubby*, which encodes a

conserved protein expressed in the hypothalamus, or in *HTR2C*, which encodes a serotonin receptor, are obese^{12,13}. Similarly, worms with a deletion in a gene encoding a key serotonin biosynthetic enzyme, *tph-1*(*mg280*)¹⁴, or in the *tubby* homologue, *tub-1*(*nr2004*)¹⁵, show an increase in fat content detectable by Nile Red. *tub-1*(*nr2004*) mutants show roughly a twofold increase in fat content, whereas *tph-1*(*mg280*) mutants accumulate about 2.5-fold more fat than wild-type animals (Supplementary Fig. 1).

To correlate the differences in body fat visualized by Nile Red with actual fat content, we extracted total lipids from wild-type and mutant animals, separated them by thin layer chromatography into triacylglyceride and phospholipid components, and quantified them by gas chromatography. Consistent with the fat histochemical assessment, *daf-2*(*e1370*) and *tph-1*(*mg280*) animals had a greater total fat content than had wild-type animals (2.5-fold and 2-fold, respectively; Fig. 1b). The increased fat contents of these animals was strongly suppressed by the *daf-16*(*mgDf47*) null mutation. In addition, excess fat in *daf-2*(*e1370*) or *tph-1*(*mg280*) animals was largely in the form of storage triacylglycerides, suggesting that in *C. elegans*, as in mammals, triacylglycerides constitute the main form of fat storage.

C. elegans genes can be inactivated by RNAi initiated with double-stranded RNA expressed in the *E. coli* food^{3,4}. A double-stranded RNAi bacterial library with 86% of the 19,000 *C. elegans* open reading frames has been constructed⁴. Analysis of this library shows that 50–90% of the genes previously identified by classical genetics can be inactivated by the clones in this library, although for genes that function in the nervous system the detection rate falls to

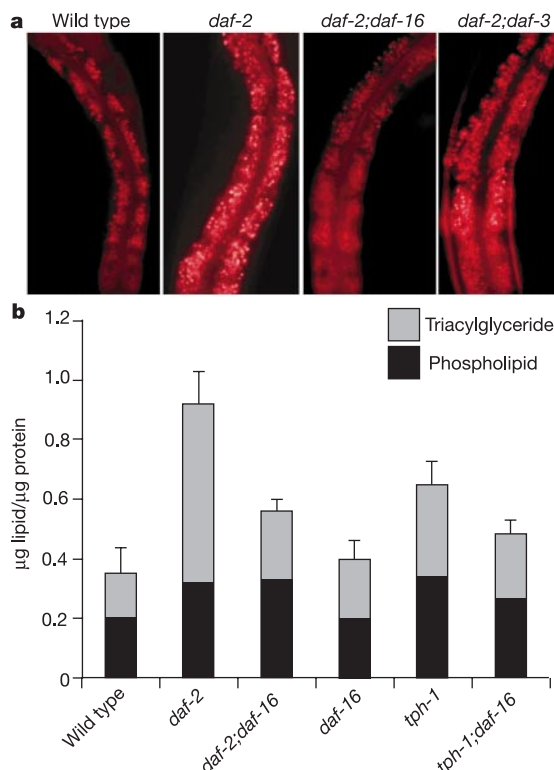


Figure 1 Changes in fat content in known mutants can be detected by Nile Red staining. **a**, Fat staining in non-starved young adult wild-type, insulin receptor mutant *daf-2*(*e1370*), *daf-16*(*mgDf47*);*daf-2*(*e1370*) and *daf-2*(*e1370*);*daf-3*(*mgDf90*) animals (rhodamine filter, original magnification $\times 160$). In each image, anterior is positioned to the bottom. **b**, Differences in Nile Red staining correspond to actual fat content measured by gas chromatography. Total lipid content, comprising triacylglycerides and phospholipids, is shown for the indicated animals. Values have been normalized to protein content extracted (refs 28, 29) from the same worms and are the means of two measurements from two independent extractions. Error bars indicate the standard deviation.

about 12% (refs 4, 16). Using the Nile Red fat assay and the RNAi library, we screened for gene inactivations that affect fat content, fat droplet morphology and the pattern of fat droplet deposition. Of 16,757 genes tested by RNAi, 1.8% (305 genes) caused reduced fat or a distorted fat deposition pattern, and 0.7% (112 genes) resulted in animals with increased fat or an enlarged fat droplet size (Tables 1 and 2 give a partial gene list; see Supplementary Tables 1 and 2 for the full gene lists. DNA and protein sequences corresponding to each *C. elegans* GenePairs designation are available at Wormbase; www.wormbase.org). RNAi inactivation of another 261 genes caused reduced fat but was also accompanied by larval arrest, embryonic lethality or sterility (data not shown). This latter group includes known genes with essential roles in fat biosynthesis and metabolism, such as acetyl-coenzyme A (CoA) carboxylase, fatty-acid synthase and the fatty-acid desaturase *fat-7*. Although the identity of these genes endorses the approach, the decreased viability makes it difficult to disentangle the fat regulatory activity of many of these 261 genes from their roles in general cellular maintenance.

The list of gene inactivations that cause altered fat content, but do not appreciably affect viability or fertility, reflects a wide range of pathways and biochemical identities. Some of these genes are already known to have a key role in mammalian fat or lipid metabolism. For example, reduced amounts of stored fat resulted from RNAi of genes encoding worm homologues of enzymatic components of the membrane lipid biosynthetic machinery (such as choline/ethanolamine phosphotransferase, and CDP-alcohol phosphatidyltransferase), β -oxidation ($\Delta 2$, $\Delta 4$, dienoylCoA reductase, 3-hydroxyacyl-CoA dehydrogenase, long chain acyl-CoA thioesterase), fatty-acid elongation enzymes and cytosolic

fatty-acid- and acyl-CoA-binding proteins. Although it is not surprising that a decrease in gene activity of fat biosynthetic enzymes would lead to a reduction in fat storage, it is less obvious why decreased gene expression of fat β -oxidation machinery would result in a similar phenotype. It is possible that inhibition of β -oxidation may produce feedback signals that alter the intake and metabolism of nutrients (see Fig. 2 for examples).

Reduced or disorganized fat deposition patterns were also caused by RNAi of several known components of sterol metabolism (such as sterol regulatory element-binding protein¹⁷), by RNAi of *C. elegans* homologues of genes that function in mammalian flux of glucose and glycerol energy metabolism (such as glyceraldehyde-3-phosphate dehydrogenase¹⁸ and phosphoenolpyruvate carboxykinase¹⁹), and by RNAi of *C. elegans* genes whose mammalian homologues are involved in gastrointestinal digestion and uptake of food (such as the oligopeptide transporter PepT1 (ref. 20) and lipases; Fig. 2).

In addition, RNAi of specific nuclear hormone receptor genes caused a reduction or an increase in fat content. Nuclear hormone receptors are known regulators of mammalian fat and sterol metabolism²¹. Notably, increased fat was caused by RNAi of a worm homologue of hepatocyte nuclear factor 4- α (HNF4- α), mutations of which are associated with maturity onset diabetes of the young²². We also observed fat phenotypes produced by RNAi of several cytochrome P450 enzymes. These enzymes may be regulated by or metabolize the ligands of the nuclear hormone receptors. The predicted *C. elegans* gene, C43H6.8, a potential orthologue of the haematopoietic/neurogenic transcription factor Nhlh-2/Nscl-2 (ref. 23), is another fat regulatory transcription factor that is conserved between mammals and *C. elegans*. Knock out of this

Table 1 RNAi clones that reduce fat in wild type and high-fat mutants

GenePairs*	Description	WT	daf-2	tph-1	tub-1	GenePairs*	Description	WT	daf-2	tph-1	tub-1
Metabolic enzyme						Transporter					
F52B11.2	Phosphomannomutase 2	Y	Y			C37A5.1	Ion channel	Y	Y		
Y55F3C.c	Thioredoxin	Y	Y			ZK682.2	Sugar transporter	Y	Y		
Y49A3A.1	CE phosphotransferase	Y	Y			K05F1.6	OCT1 organic carrier	Y	Y		
C06E7.3	S-adenosylmethionine synthase	Y	Y	Y	Y	C13D9.7	Na ⁺ /Ca ²⁺ exchanger	Y	Y		Y
C36A4.9	Acetyl-CoA synthetase	Y	Y	Y	Y	C32C4.1	Potassium channel	Y	Y		Y
F08F8.2	HMG-CoA reductase	Y	Y	Y	Y	F15H10.4	Lysosomal transporter	Y	Y	Y	Y
K10B3.7	GAPDH	Y	Y	Y	Y	K04E7.2	PepT1 symporter	Y	Y	Y	Y
F11E6.5	Fatty acid elongase	Y	Y	Y	Y	C34G6.4	ABC transporter	Y	Y	Y	Y
F13D11.1	Lysosomal acid phosphatase	Y	Y	Y	Y	Energy metabolism					
K07C6.5	Cytochrome P450 2C2	Y	Y	Y	Y	F20D1.9	Uncoupling protein	Y	Y		
K09D9.2	Cytochrome P450	Y	Y	Y	Y	F28H6.2	Energy transfer protein	Y	Y		Y
Fat/lipid interacting						Vesicular transport					
R07B7.9	Brush border esterase/lipase	Y	Y			F11A5.3	RAB2	Y	Y		
F13D12.6	Esterase/lipase	Y	Y	Y	Y	R01H2.3	Sortilin	Y	Y	Y	Y
Transcription factor						F53H8.1	Clathrin adaptor chain	Y	Y	Y	Y
ZK686.4	Zinc-finger C ₂ H ₂ type	Y	Y			Protein degradation					
F11A1.3	NHR (dihydroxyvitamin D3)	Y	Y			Y65B4B10e	Ubiquitin protein ligase	Y	Y		
T09F3.1	Zinc-finger C ₂ H ₂ type	Y	Y			F49E12.4	Ubiquitin-conjugating enzyme	Y	Y	Y	
K08A2.b	NHR (HNF-4 α)	Y	Y			C49C3.3	Ubiquitin family	Y	Y	Y	Y
F22A3.5	Pre-B-cell leukaemia factor	Y	Y		Y	Other					
Y47D3B.7	SREBP	Y	Y	Y	Y	Y77E11A	Collagen	Y	Y		
Signal transduction						C06G3.2	Kinesin motor domain	Y	Y	Y	
C33H5.17	D111/G-patch domain	Y	Y			R04A9.4	Translation initiation	Y	Y	Y	
T04D3.2	EF-hand family domain	Y	Y			No function assigned					
ZC302.1	Serine/threonine protein phosphatase	Y	Y			Y24D9A.b	Unknown	Y	Y		
ZC504.4	Tyrosine and serine/threonine kinase	Y	Y			B0034.2	Unknown	Y	Y		
T19D2.2	Dual specificity phosphatase	Y	Y	Y		C30G4.5	Unknown	Y	Y		
F46G11.3	Protein kinase	Y	Y	Y	Y	T10D4.1	Unknown	Y	Y		
M01B12.5	Tyrosine kinase	Y	Y	Y	Y	R08F11.2	Unknown	Y	Y		Y
ZK675.1	Patched	Y	Y	Y	Y	K12D12.4	Unknown	Y	Y		
Receptor						B0554.7	Unknown	Y	Y		
H09F14.1	GPCR	Y	Y			F46F5.10	Unknown	Y	Y		
C38C10.1	Neurokinin-type receptor	Y	Y			F10A3.11	Unknown	Y	Y		
F07C4.1	GPCR	Y	Y	Y	Y	B0554.6	Unknown	Y	Y	Y	
T07C12.5	GPCR	Y	Y	Y	Y	T01D3.4	Unknown	Y	Y		Y
T14E8.3	Dopamine receptor	Y	Y	Y	Y	C18E9.5	Unknown	Y	Y		Y
Y40H7A.7	SRA family chemoreceptor	Y	Y	Y	Y	K09H11.2	Unknown	Y	Y	Y	Y
Neuronal						ZK131.8	Unknown	Y	Y	Y	Y
H27A22.1	Glutamyl cyclase	Y	Y			F52C6.12	Unknown	Y	Y	Y	Y
T19B4.7	DCC/axon guidance	Y	Y			F59F5.2	Unknown	Y	Y	Y	Y

Clones were tested on wild-type (WT), *daf-2(e1370)*, *tph-1(mg280)* and *tub-1(nr2004)* animals. 'Y' indicates that RNAi caused a reduced fat phenotype in the fat content of the test strain. ABC, ATP-binding cassette. NHR, nuclear hormone receptor; GPCR, G-protein-coupled receptor; DCC, deleted in colorectal cancer.

transcription factor causes obesity in mice²³ and RNAi of C43H6.8 resulted in an increase in fat content (Fig. 2).

RNAi of several *C. elegans* genes that may function in food sensation and neuroendocrine signalling also resulted in aberrant fat content. RNAi of a likely glutamate receptor and of a protein with homology to rat hippocampal somatostatin receptor led to increased fat storage. Reduced fat was caused by RNAi of the potential orthologue of glutamyl cyclase (required for biosynthesis of pyroglutamyl peptides), a dopamine D2-like receptor, and several chemoreceptors and olfactory receptors. The list of genes that regulate fat is unexpectedly diverse. A future challenge for genome-wide functional studies is the organization and prioritization of the identified genes. It is likely, for example, that both signal transduction components and their targets have been identified in this screen.

To organize the newly identified fat regulatory genes into signalling pathways, we tested the effect of each of the 305 gene inactivations on three *C. elegans* mutant strains with neuroendocrine signalling defects that result in increased fat: the *daf-2* insulin receptor mutant⁶, the *tph-1* serotonin deficient mutant¹⁴, and the *tub-1* tubby orthologue mutant¹⁵. Of the 305 RNAi clones tested, 32 (11%) caused substantial reductions in fat in wild type and in each of the three obese strains (Table 1). This group may represent the key fat regulatory genes that, when inactivated, potently and specifically affect fat storage without obvious side-effects on the general health of the animal. A third of the genes on this list correspond to metabolic enzymes and other well-characterized fat homeostasis genes. But many receptors, signal transduction molecules, transporters and vesicular transport molecules are also present on this list.

The transporters and signal transduction components may constitute central factors in the transport of fat, or in the feedback signalling from fat storage tissues to neuronal regulators of feeding and metabolism. Similarly, the vesicular sorting and protein degradation components may be the molecules that regulate the cellular trafficking of fat/metabolite transporters and receptors. These findings endorse the notion that body weight is the integration of several inputs, even in a relatively simple organism like *C. elegans*. Notably, most of these *C. elegans* fat regulatory genes have mammalian homologues.

Another set of RNAi clones showed a distinct pattern of reversing the high fat phenotype: numerous RNAi clones that potently reduced fat content in wild-type or *daf-2* animals failed to alter

the increased fat content of *tph-1* and *tub-1* mutant animals (Table 1). There was a strong correlation (86%) between the effectiveness or failure of these RNAi clones to modulate fat quantities in either *tph-1* or *tub-1* mutant animals. Failure of an RNAi clone in this group to generate fat reduction in either the *tph-1* or *tub-1* mutant could not be attributed to weak or partial RNAi activity, because RNAi clones in this group were effective against *daf-2* animals, which had the most fat among the strains tested. These results indicate that the fat-increasing signals caused by the *tph-1* and *tub-1* mutations share at least one common pathway that is separable from the fat-increasing signal caused by a *daf-2* mutation. This notion is supported by the fact that both worm serotonin¹⁴ and tubby signalling pathways act in the nervous system (unpublished observations), whereas the fat regulatory activity of the insulin like pathway is not limited to the nervous system^{24,25}.

By contrast, another set of RNAi clones reduced fat in wild type, *tph-1* or *tub-1* animals, but not in *daf-2* mutant animals. Notably, this list was enriched in signal transduction molecules and receptors (Supplementary Table 3). We suggest that the vesicular transporters, receptors, and signal transduction molecules on this list are candidates for transmitting fat regulatory signals from tissues affected by the *tph-1* and *tub-1* mutations. Analogous roles in neuroendocrine control of body weight regulation for mammalian homologues of these genes would be expected. Finally, 52% of the fat-reducing RNAi clones caused fat reduction only in wild type, but did not affect the greater fat content of *daf-2*, *tph-1* and *tub-1* mutants (Supplementary Table 1). These genes may have modulatory roles in the regulation of organismal fat.

The diverse array of the genes identified in our screen provides an unprecedented glimpse of the cellular machineries that regulate body fat content and deposition. To highlight the shared ancestry of *C. elegans* and mammalian fat storage regulation, we could identify *C. elegans* genes for which the mammalian homologues have already been implicated in body weight content. The concordance between the list of *C. elegans* and mammalian fat regulatory and obesity genes suggests that these animals show ancient and common pathways of body fat regulation.

Notably, over 50% of the *C. elegans* fat regulatory genes identified in our screen have mammalian homologues that have not been previously implicated in regulating fat storage. It is possible that homologues of these newly identified *C. elegans* fat regulatory genes also control mammalian body weight. Their contributions to

Table 2 Partial list of RNAi targets that produce increased fat content

GenePairs*	Description	GenePairs*	Description
Metabolic enzyme (7)		Signal transduction (15)	
E04F6.3	Hydroxysteroid 17-β dehydrogenase	K08F8.1	Ribosomal S6 kinase
F28F8.2	Long-chain fatty-acid CoA ligase	F56H11.6	Casein/tau-tubulin kinase
ZC513.1	Phospholipid transfer protein	C04G2.2	Serine/threonine kinase
C33A12.6	UDP-glucosyl transferase	F39B1.1	Phosphoinositide 3-kinase
VF13D12L.1	Myo-inositol-1-phosphate synthase	W08D2.1	Wnt-1 family kinase
Transcription factors (9)		W03A5.4	Guanylate kinase associated
C43H6.8	Nhlh2/Nscl-2	C24F3.2	Glucokinase-associated, phosphatase
K10C3.6	NHF-4 α subtype NHR	F46C5.6	Protein phosphatase PP2A
F33D4.1	Oestrogen-type NHR	T04C9.1	Oligophrenin-1 GTPase
C56C10.10	Aryl hydrocarbon NHR	Y11D7A.9	FGF receptor activating protein
C37F5.1	Elk-1	C18H9.7	RAPSN
R11H6.5	Interleukin enhancer factor 2	Channels/transporters (4)	
H12C20.3	C4-type steroid receptor	ZC410.4	Potassium channel
Receptors (6)		F52H2.2	Amino acid permease
F56B6.5	Somatostatin receptor-type	F14E5.1	Glucose transporter-3
C43H6.9	Glutamate receptor	Cell surface/structural (5)	
Y27F2A.g	Chemoreceptor	C34F6.3	Collagen triple helix repeat
Y46H3C_11.b	GPCR	Y38F1A.9	Contactin 6/myopalladin
Vesicular transport (3)		No function assigned (62)	
F08H9.5	Cubilin/endocytic receptor		
C04G2.4	Vesicle associated protein		

*For each RNAi clone the Research Genetics GenePairs designation is provided. Genes corresponding to the RNAi clones were grouped into functional classes. Total number of genes assigned to each functional category appear in parenthesis. A full list including the extent of increase in fat content is given in the Supplementary Information. FGF, fibroblast, growth factor. NHR, nuclear hormone receptor; GPCR, G-protein-coupled receptor; RAPSN, acetylcholine receptor-associated protein.

mammalian body fat regulation can be studied in tissue culture or in rodent models. Given the complex interactions between fat cells and the central nervous system^{1,2}, however, the study of fat regulation in a physiologically intact animal, even an animal as distant from mammals as *C. elegans*, can provide insights that are not possible from studying, for example, isolated mammalian adipocytes.

Indeed, an important aspect of body weight regulation in both mammals and *C. elegans* is the profound link between feeding and body fat content in both systems. Despite the fact that *C. elegans* lacks dedicated adipocytes and sequence identifiable leptin, worm body fat is likely to be coupled to food sensation. In addition to the known impact of insulin and serotonin signalling, our results suggest the involvement of dopaminergic and glutamate signalling in regulating body fat. The G-protein-coupled receptors (GPCRs) detected in our RNAi analysis may lead to identification of neuropeptide and other neuroendocrine signalling pathways that regulate fat content and feeding behaviour. Further studies of the rate of feeding in animals with deficits in these receptors might distinguish whether the affected pathways regulate food intake behaviour or metabolism.

Many other *C. elegans* genes that regulate body fat remain to be identified. Not all genes are responsive to RNAi inactivation. Some genes, most notably neuronally expressed genes, may be only partially inactivated by RNAi. The current RNAi library covers about 86% of the known and predicted genes, and certain classes of gene may be missing or underrepresented in it. In addition, some gene inactivations probably manifest phenotypes only under specific environmental conditions or in specific sensitised genetic backgrounds.

A genome-wide RNAi screen combines the capacity of genetics to identify key components of pathways, regardless of transcriptional status, with the comprehensiveness and facility of transcriptional profiling studies. Unlike the list of genes generated from transcriptional profiling studies, a genetic function in *C. elegans* is already established for the list of genes generated by this genome-wide RNAi screen. But a useful way to prioritize our list of fat regulatory genes will be to compare it with the lists of transcriptionally regulated genes in various models of obesity.

An immediate application of the study of *C. elegans* fat regulation will be to pedigree analysis in human obesity and lipodystrophy syndromes and to quantitative trait studies in rodents. Genes discovered in *C. elegans* can point to candidate obesity or diabetes loci in the broad chromosomal regions identified in human or rodent studies²⁶. Our list of *C. elegans* fat regulatory genes identifies about 150 of the 30,000 mammalian genes that might be tested for variants in obese or diabetic populations. About 20% of the *C. elegans* genes identified in this screen caused an increase in fat when inactivated, suggesting that loss-of-function mutations in their mammalian homologues could underlie obesity. But most of the fat regulatory genes that we identified caused a decrease in fat storage and thus would be expected to cause mammalian obesity only if activated by mutation. Because inactivation of some of these genes diminishes fat content even in *C. elegans* mutants with defects in insulin, serotonin and tubby signalling pathways, these genes are promising candidates for developing drugs to treat obesity and its associated diseases. It is significant that many of the genes encode enzymes or proteins that are members of families with known small-molecule regulators (such as nuclear hormone receptors, potassium channels, neurotransmitter receptors and kinases), or small active sites that are good candidates for drug development. □

Received 27 August; accepted 29 October 2002; doi:10.1038/nature01279.

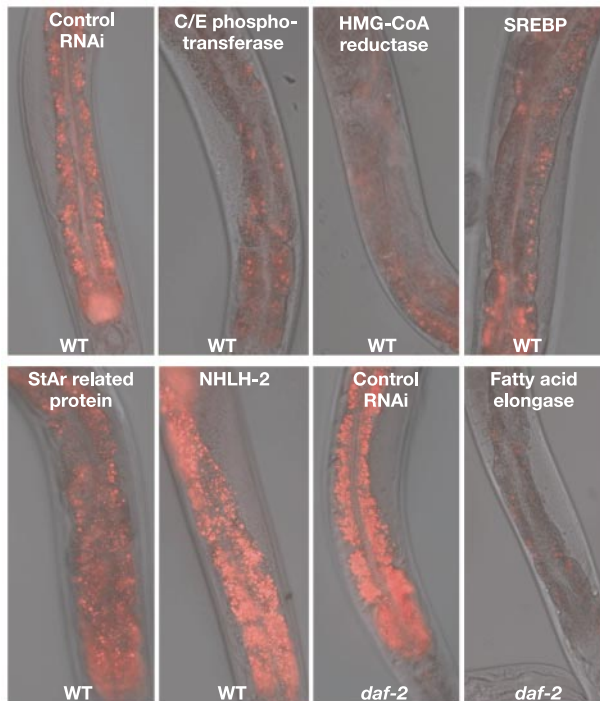


Figure 2 Examples of RNAi targets that produce alterations in fat content. Wild-type and *daf-2(e1370)* animals stained with Nile Red are shown. Reduced fat staining of wild-type animals was caused by RNAi of Y49A3A.1 (choline/ethanolamine phosphotransferase), F08F8.2 (hydroxymethylglutaryl-CoA reductase) and Y47D3B.7 (SREBP). Distorted fat deposition pattern was caused by RNAi of K02D3.2 (steroidogenic acute regulatory related protein). Increased fat was caused by RNAi of C43H6.8 (Nhlh2 transcription factor). RNAi of F11E6.5 (fatty-acid elongase) caused reduced fat in *daf-2(1370)* animals. An L4440 vector was used for control RNAi. Overlays of Nomarski and rhodamine images are shown. In each image, anterior is positioned to the bottom.

- Schwartz, M. W., Woods, S. C., Porte, D. Jr, Seeley, R. J. & Baskin, D. G. Central nervous system control of food intake. *Nature* **404**, 661–671 (2000).
- Friedman, J. M. & Halaas, J. L. Leptin and the regulation of body weight in mammals. *Nature* **395**, 763–770 (1998).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
- Greenspan, P. & Fowler, S. D. Spectrofluorometric studies of the lipid probe, Nile Red. *J. Lipid Res.* **26**, 781–789 (1985).
- Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* **277**, 942–946 (1997).
- Riddle, D. L. & Albert, P. S. *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 739–768 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1997).
- Wadsworth, W. G. & Riddle, D. L. Developmental regulation of energy metabolism in *Caenorhabditis elegans*. *Dev. Biol.* **132**, 167–173 (1989).
- Ogg, S. *et al.* The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389**, 994–999 (1997).
- Lin, K. *et al.* An HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. *Science* **278**, 1319–1322 (1997).
- Ogg, S. & Ruvkun, G. The *C. elegans* PTEN homolog, DAF-18, acts in the insulin receptor-like metabolic signaling pathway. *Mol. Cell* **2**, 887–893 (1998).
- Tecott, L. H. *et al.* Eating disorder and epilepsy in mice lacking 5-HT_{2c} serotonin receptors. *Nature* **374**, 542–546 (1995).
- Noben-Trauth, K., Naggert, J. K., North, M. A. & Nishina, P. M. A candidate gene for the mouse mutation *tubby*. *Nature* **380**, 534–538 (1996).
- Sze, J. Y., Victor, M., Loer, C., Shi, Y. & Ruvkun, G. Food and metabolic signalling defects in a *Caenorhabditis elegans* serotonin-synthesis mutant. *Nature* **403**, 560–564 (2000).
- Liu, L. X. *et al.* High-throughput isolation of *Caenorhabditis elegans* deletion mutants. *Genome Res.* **9**, 859–867 (1999).
- Kamath, R. S., Martinez-Campos, M., Zipperlen, P., Fraser, A. G. & Ahringer, J. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biol.* **2**, 1–10 (2000).
- Brown, M. S. & Goldstein, J. L. A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood. *Proc. Natl Acad. Sci. USA* **96**, 11041–11048 (1999).
- Rolland, V. *et al.* Evidence of increased glyceraldehyde-3-phosphate dehydrogenase and fatty acid synthetase promoter activities in transiently transfected adipocytes from genetically obese rats. *J. Biol. Chem.* **270**, 1102–1106 (1995).
- Olswang, Y. *et al.* A mutation in the peroxisome proliferator-activated receptor γ binding site in the gene for the cytosolic form of phosphoenolpyruvate carboxykinase reduces adipose tissue size and fat content in mice. *Proc. Natl Acad. Sci. USA* **99**, 625–630 (2002).
- Adibi, S. A. The oligopeptide transporter (Pept-1) in human intestine: biology and function. *Gastroenterology* **113**, 332–340 (1997).

21. Chawla, A., Repa, J. J., Evans, R. M. & Mangelsdorf, D. J. Nuclear receptors and lipid physiology: opening the X-files. *Science* **294**, 1866–1870 (2001).
22. Yamagata, K. *et al.* Mutations in the hepatocyte nuclear factor-4 α gene in maturity-onset diabetes of the young (MODY1). *Nature* **384**, 458–460 (1996).
23. Good, D. J. *et al.* Hypogonadism and obesity in mice with a targeted deletion of the *Nhlh2* gene. *Nature Genet* **15**, 397–401 (1997).
24. Apfeld, J. & Kenyon, C. Regulation of lifespan by sensory perception in *Caenorhabditis elegans*. *Nature* **402**, 804–809 (1999).
25. Wolkow, C. A., Kimura, K. D., Lee, M. S. & Ruvkun, G. Regulation of *C. elegans* life-span by insulin-like signaling in the nervous system. *Science* **290**, 147–150 (2000).
26. Brockmann, G. A. & Bevoval, M. R. Using mouse models to dissect the genetics of obesity. *Trends Genet.* **18**, 367–376 (2002).
27. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
28. Johnstone, I. L. *C. elegans: A Practical Approach* (ed. Hope, I. A.) 201–225 (Oxford Univ. Press, Oxford, 1999).
29. Watts, J. L. & Browse, J. Genetic dissection of polyunsaturated fatty acid synthesis in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **99**, 5854–5859 (2002).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank M. Basson and C. Johnson for the *tub-1(nr2004)* strain. K.A. is grateful to A. Pasquinelli and other members of the Ruvkun laboratory for discussions and suggestions. K.A. was supported by a fellowship from the Damon-Runyon Walter-Winchell Cancer Research Fund; J.W. by a NIH grant; A.G.F. by a US Army Breast Cancer Research Fellowship; R.S.K. by a Howard Hughes Medical Institute Predoctoral Fellowship; and J.A. by a Wellcome Trust Senior Research Fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.R.
(e-mail: ruvkun@molbio.mgh.harvard.edu).

Coupling of agonist binding to channel gating in the GABA_A receptor

Thomas L. Kash^{*}, Andrew Jenkins[†], Jill C. Kelley[†], James R. Trudell[‡] & Neil L. Harrison[†]

^{*} Graduate Program in Neuroscience, Weill Graduate School of Biomedical Sciences, Cornell University, New York, New York 10021, USA

[†] Departments of Anesthesiology and Pharmacology, Weill Medical College, Cornell University, New York, New York 10021, USA

[‡] Department of Anesthesia and Beckman Program for Molecular and Genetic Medicine, Stanford University, Stanford, California 94305-5117, USA

Neurotransmitters such as acetylcholine and GABA (γ -aminobutyric acid) mediate rapid synaptic transmission by activating receptors belonging to the gene superfamily of ligand-gated ion channels (LGICs)¹. These channels are pentameric proteins that function as signal transducers, converting chemical messages into electrical signals². Neurotransmitters activate LGICs by interacting with a ligand-binding site^{3–7}, triggering a conformational change in the protein that results in the opening of an ion channel⁸. This process, which is known as 'gating', occurs rapidly and reversibly, but the molecular rearrangements involved are not well understood⁹. Here we show that optimal gating in the GABA_A receptor, a member of the LGIC superfamily, is dependent on electrostatic interactions between the negatively charged Asp 57 and Asp 149 residues in extracellular loops 2 and 7, and the positively charged Lys 279 residue in the transmembrane 2–3 linker region of the α_1 -subunit. During gating, Asp 149 and Lys 279 seem to move closer to one another, providing a potential mechanism for the coupling of ligand binding to opening of the ion channel.

Many inherited mutations in the transmembrane 2–3 linker (2–3L) region of LGICs alter gating and are associated with human

diseases: defects in the nicotinic acetylcholine receptor (nAChR) give rise to forms of myasthenic syndrome¹⁰, defects in the glycine receptor are associated with hyperekplexia¹¹, and defects in the GABA_A receptor (GABA_A-R) are associated with rare forms of epilepsy¹². The clustering of gating mutations in the 2–3L region has led to the suggestion that this loop is essential for communicating the conformational changes that result from interaction of neurotransmitter with the ligand-binding site to the transmembrane domain and its integral ion channel^{9,13}.

An acetylcholine-binding protein (AChBP) in the snail *Lymnaea stagnalis* has been found to have a high degree of homology with the extracellular domains of LGIC subunits, and its structure has been solved at high resolution⁴. The location of two flexible loops (loops 2 and 7) in this structure is striking. As pointed out by others⁹, "in the full receptor, the signature loop is positioned to interact directly with the membrane, or possibly with the transmembrane regions (or the short sequence connecting two of them) of the receptor. The implication is that the signature loop might be involved in 'gating'". In the LGIC subunits, the 'signature loop' (loop 7) is formed between two cysteine residues that are joined by a disulphide bond, which is an essential feature of these proteins. We also reasoned that homologous loops in the GABA_A-R would be positioned close to the interface between the extracellular domain and the membrane, enabling loops 2 and 7 to interact with the 2–3L region.

We aligned the amino acid sequences of the GABA_A-R α_1 subunit and the AChBP (see Supplementary Information) and found that loops 2 and 7 in the GABA_A-R α_1 -subunits contain several highly conserved acidic residues, whereas the 2–3L region contains two highly conserved basic residues. We therefore proposed the hypothesis that electrostatic interactions between one or more residues in loops 2 and 7 and those in the 2–3L region contribute to the gating mechanism in the GABA_A-R. This hypothesis was tested by site-

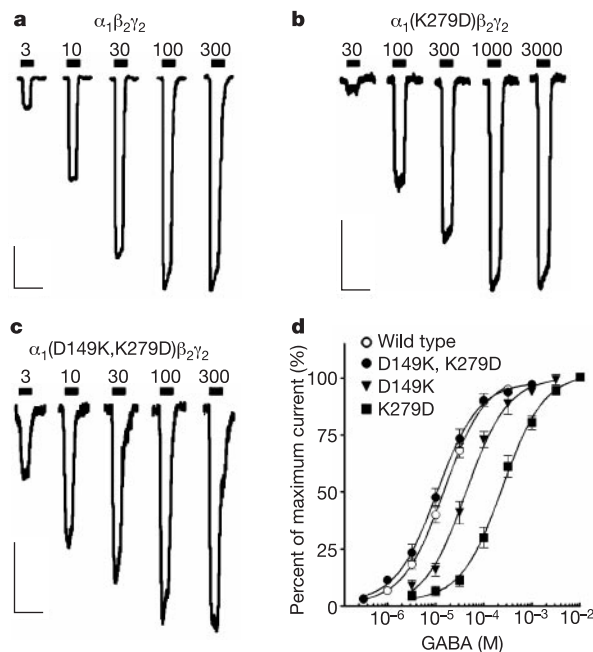


Figure 1 Charge exchange between positions 149 and 279 restores optimal GABA sensitivity. **a–c**, Representative recordings from HEK 293 cells transfected with $\alpha_1\beta_2\gamma_2$ s (**a**), α_1 (K279D) $\beta_2\gamma_2$ s (**b**) and α_1 (D149K) $\beta_2\gamma_2$ s (**c**) receptors. The bars above the recordings indicate the durations of GABA application, and the numbers above each bar indicate the concentration of applied GABA in micromoles. Scale bars, 200 pA and 10 s. **d**, Concentration–response curves for $\alpha_1\beta_2\gamma_2$ s (open circles), α_1 (D149K) $\beta_2\gamma_2$ s (filled triangles), α_1 (K279D) $\beta_2\gamma_2$ s (filled squares) and α_1 (D149K, K279D) $\beta_2\gamma_2$ s (filled circles).

Table 1 Properties of selected mutant receptors

Receptor	EC ₅₀ (GABA, μ M)	n_H (GABA)	$I_{\max}$ (pA) (GABA)	EC ₅₀ (P4S, μ M)	n_H (P4S)	ε	$\Delta\Delta G$ (kJ mol ⁻¹)
Wild type	21 $\pm$ 2 (59)	1.33 $\pm$ 0.07 (59)	-530 (59)	27 $\pm$ 11 (9)	1.2 $\pm$ 0.1 (9)	0.76 $\pm$ 0.04 (9)	
α_1 (K279D) $\beta_2\gamma_{2s}$	205 $\pm$ 48*** (15)	1.34 $\pm$ 0.16 (15)	-409 (15)	235 $\pm$ 48 (9)	0.7 $\pm$ 0.1 (9)	0.20 $\pm$ 0.05*** (9)	5.7 $\pm$ 1.5
α_1 (D149K) $\beta_2\gamma_{2s}$	48 $\pm$ 8** (11)	1.16 $\pm$ 0.09 (11)	-175 (11)	100 $\pm$ 26 (6)	1.3 $\pm$ 0.5 (6)	0.37 $\pm$ 0.06** (6)	2.0 $\pm$ 0.4
α_1 (D57K) $\beta_2\gamma_{2s}$	32 $\pm$ 3** (6)	1.16 $\pm$ 0.11 (6)	-681 (6)	25 $\pm$ 7 (5)	1.0 $\pm$ 0.1 (5)	0.61 $\pm$ 0.04* (5)	1.1 $\pm$ 0.1
α_1 (D149K,K279D) $\beta_2\gamma_{2s}$	16 $\pm$ 3††† (21)	1.30 $\pm$ 0.20 (21)	-202 (21)	109 $\pm$ 26 (9)	0.8 $\pm$ 0.1 (9)	0.69 $\pm$ 0.05††† (9)	-0.7 $\pm$ 0.1
α_1 (D57K,K279D) $\beta_2\gamma_{2s}$	16 $\pm$ 2††† (6)	1.50 $\pm$ 0.06 (6)	-179 (6)	280 $\pm$ 20 (5)	1.3 $\pm$ 0.2 (5)	0.74 $\pm$ 0.06††† (5)	-0.7 $\pm$ 0.1

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ as compared with wild type; ††† $P < 0.001$ as compared with α_1 (K279D) $\beta_2\gamma_{2s}$

directed mutagenesis using a combination of charge reversal, charge exchange and cysteine crosslinking approaches.

Previous studies have implicated K279 in the 2–3L region in gating^{13,14}. To determine the effect of charge reversal, receptors containing a mutant α_1 (K279D) subunit were expressed and compared with wild-type receptors (Fig. 1a, b). This mutation caused a significant decrease in the GABA sensitivity of the receptor (Fig. 1b, d, and Table 1). In addition, piperidine-4-sulphonic acid (P4S), a partial agonist of the GABA_A-R^{13,15,16}, produced maximal responses in α_1 (K279D) $\beta_2\gamma_{2s}$ receptors that were on average about 20% of the amplitude of the maximal current elicited by GABA (Table 1), showing that the relative efficacy (ε) of P4S is decreased in the mutant α_1 (K279D) $\beta_2\gamma_{2s}$ receptor relative to the wild-type receptor. Such decreases in agonist efficacy are characteristic of deleterious gating mutations.

The following charge reversal mutants were also made for the loop 2 and loop 7 residues: α_1 (D55K), α_1 (D57K), α_1 (E59K), α_1 (D145K) and α_1 (D149K). One of these mutations, α_1 (D55K), produced a receptor that failed to show any significant current response to GABA. There was a significant decrease in the sensitivity to GABA in the mutant GABA_A-Rs incorporating α_1 (D57K), α_1 (E59K) and α_1 (D149K), but not in α_1 (D145K) (Table 1 and Supplementary Information). In addition, α_1 (D57K), and α_1 (D149K) mutations caused a significant decrease in the relative efficacy of the partial agonist (Table 1), again consistent with the idea that these mutations influence gating. The decrease in GABA sensitivity of the mutant receptors can be ascribed to an increase in the free-energy change (ΔG) for activation. The $\Delta\Delta G$ values associated with each mutation are listed in Table 1 (see also Supplementary Information).

To address the possibility that residues in loops 2 and 7 interact closely with the 2–3L region, we created a series of double ‘charge exchange’ mutants: for loop 7, α_1 (D145K,K279D) and α_1 (D149K,K279D); and for loop 2, α_1 (D57K, K279D) and α_1 (E59K, K279D). We considered that if acidic and basic residues were interacting in the wild-type GABA_A-R then this exchange of charges might restore normal function. The defect associated with the α_1 (K279D) mutation was indeed corrected in the double mutants α_1 (D149K, K279D) and α_1 (D57K, K279D) (Fig 1c, d; and Table 1). In addition, the relative efficacy of the partial agonist at the double-mutant receptors was indistinguishable from that of wild type (Table 1).

We investigated further the interaction between D149 and K279 by mutating both residues to the neutral hydrophobic residue alanine, both individually and in combination. The single mutations D149A and K279A produced decreases in GABA sensitivity similar to the single charge reversals, whereas the double-mutant α_1 (D149A, K279A) failed to show any significant current in response to GABA (see Supplementary Information). It therefore seems that electrostatic interactions between D149 and K279 are necessary for optimal receptor gating, and that these cannot be replaced by simple hydrophobic interactions between alanine residues.

The technique of mutant cycle analysis has been used extensively to calculate coupling energies between amino acid residues intimately involved in protein–protein interactions^{17,18}. We adopted

this concept to investigate the significance of the intermolecular interactions between the charged residues. The coupling coefficient (Ω)¹⁸ for the D57K–K279D interaction predicts a coupling energy of 7.7 kJ mol⁻¹ (Fig. 2a), whereas the value for D145K–K279D is close to zero (Fig. 2b). A coupling energy of 7 kJ mol⁻¹ is characteristic of charged residues separated by 5 Å or less¹⁹.

We also used a cysteine crosslinking technique^{20,21} to confirm the proximity between D149 in loop 7, D57 in loop 2, and K279. In proteins of known structure, the average separation between the α -carbons of disulphide-linked cysteine residues is 5.6 Å (ref. 22). We reasoned that it should therefore be possible to detect cross-linking between cysteines engineered at positions 149 and/or 57 with a cysteine introduced at position 279. In the wild-type receptor, application of the oxidizing agent Cu:phenanthroline

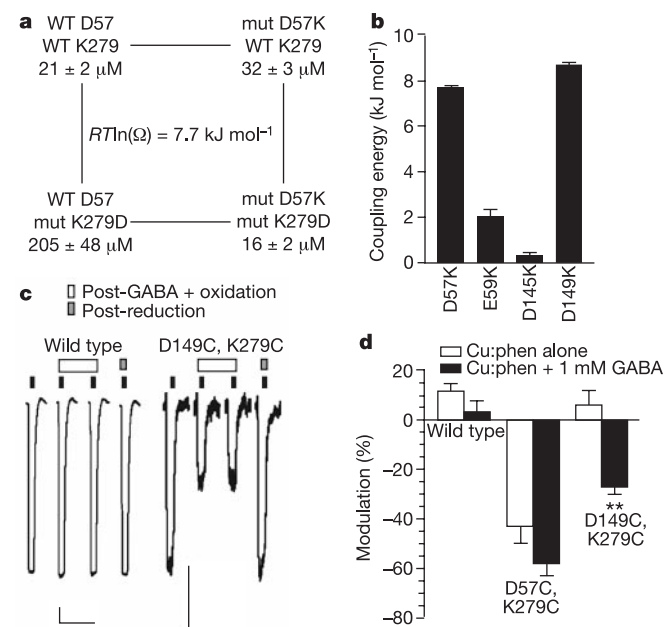


Figure 2 Mutant cycle analysis and disulphide crosslinking experiments support interactions between loops 2 and 7 and the 2–3L region. **a**, Mutant cycle analysis showing strong coupling between D57 and K279. **b**, Histogram showing the calculated coupling energy (kJ mol⁻¹) between each of the four extracellular residues and K279 in the 2–3L region. **c**, A 2-min application of 1 mM GABA and Cu:phen has no effect on responses to subsequent applications of GABA (20 μ M) in wild-type $\alpha_1\beta_2\gamma_{2s}$ receptors. Black bars indicate the duration of GABA application, open bars highlight currents that were recorded after a 2-min application of 1 mM GABA and Cu:phen, and grey bars highlight currents recorded after the application of DTT. Conversely, a 2-min application of 1 mM GABA and Cu:phen inhibits responses to subsequent applications of GABA (30 μ M) in mutant α_1 (D149C,K279C) $\beta_2\gamma_{2s}$ receptors. The inhibition is reversed by application of 10 mM DTT. Scale bars, 100 pA and 50 s (wild type); 15 pA and 50 s (D149C, K279C). **d**, Change in receptor function measured by response to an EC₅₀ concentration of GABA after treatment with oxidizing reagent alone (Cu:phen), or oxidizing reagent plus maximal agonist (Cu:phen + 1 mM GABA) on wild-type, α_1 (D57K,K279C) $\beta_2\gamma_{2s}$ and α_1 (D149C,K279C) $\beta_2\gamma_{2s}$ receptors. The modulation of α_1 (D149C,K279C) $\beta_2\gamma_{2s}$ by Cu:phen and GABA is significantly different from the effect of Cu:phen alone (** $P < 0.01$).

(Cu:phen) or the reducing agent dithiothreitol (DTT) had no effect on receptor function, whether the reagents were applied in the absence or presence of GABA (Fig. 2c, d). Similarly, the single-mutant receptors incorporating α_1 (D57C), α_1 (D149C) or α_1 (K279C) were also unaffected by Cu:phen or DTT (data not shown).

By contrast, in the α_1 (D149C,K279C) double mutant (Fig. 2c), simultaneous treatment with GABA and Cu:phen produced a significant decrease in the amplitude of currents induced by subsequent applications of GABA (applied at the effective concentration for half-maximum response (EC_{50})). This effect was reversed following application of DTT. No effect of Cu:phen was observed in the absence of GABA, even when Cu:phen was applied for up to 10 min (Fig. 2d). In the α_1 (D57C,K279C) double mutant, Cu:phen caused a decrease in receptor function (Fig. 2d), without any measurable difference in the rate of the reaction in the absence or presence of GABA (data not shown). These results are consistent with the formation of a covalent linkage between the cysteines engineered into the GABA_A-R at positions 57 and 279, or at 149 and 279. In the case of positions 149 and 279, this reaction is dependent on the presence of a saturating concentration of GABA. The formation of disulphide bonds between these engineered cysteines provides complementary evidence for the close proximity (~ 5 Å) of both D57 and D149 to K279.

We found that charge exchange between K279 and either of two of the negatively charged residues in loops 2 and 7 conserved optimal GABA_A-R function, indicating that electrostatic interactions may occur between K279 in the 2–3L region and the negatively charged residues of loops 2 and 7. This concept is shown in a molecular model (Fig. 3a), in which the extracellular domain of the GABA_A-R α -subunit (modelled based on the structure of the AChBP) is combined with a model of the transmembrane domain of the GABA_A-R²³. The model reveals the proximity of loops 2 and 7 to the 2–3L. Our results suggest that interactions occur between K279 and two negatively charged residues, and this idea is clearly shown in our molecular model (Fig. 3b).

One of the most interesting observations is the finding that disulphide bond formation between positions 149 and 279 takes

place only in the presence of a high concentration of GABA (whereas it occurred independently of GABA for positions 57 and 279). The consequence of disulphide bond formation in the α_1 (D149C,K279C) double mutant was to inhibit the response to subsequent applications of GABA. One interpretation of this observation is that crosslinking traps the receptor in the desensitized state, a suggestion supported by the fact that treatment with a lower (non-desensitizing) concentration of GABA in the presence of Cu:phen had no effect (data not shown). The state-dependence of this reaction is consistent with the idea that D149 and K279 move closer together during the gating process, and that the relative movements of loop 7 and the 2–3L region may be a crucial component of the structural rearrangements associated with gating.

We propose that specific electrostatic interactions provide an intramolecular coupling mechanism in the α -subunit of the GABA_A-R. Such interactions are energetically well suited to participate in the rapid molecular rearrangements associated with gating^{8,16}. A marked difference between the GABA_A-R and nAChR sequences is the lack of a charge pair in the nAChR analogous to D149–K279. Inspection of the nAChR sequences shows that there are conserved serine and threonine residues in these loops, so it is possible that, in the nAChR, hydrogen bonding takes the place of electrostatic interactions. Structural work on the nAChR of *Torpedo*²⁴ shows that a main consequence of agonist binding to the receptor is the relative movement of parts of the extracellular domain, and interactions between loop 7 and the 2–3L region have been proposed to result from such a conformational change^{9,24}. The experimental results that we describe here for the GABA_A-R are clearly consistent with this suggestion—relative movement between the signature loop and the 2–3L region does indeed seem likely to underlie gating in at least one member of the LGIC superfamily.

Note added in proof: A recent study²⁸ shows that the accessibility of residues within the transmembrane 2–3 linker of the GABA_A receptor α -subunit is altered in the presence of GABA, consistent with a conformational change in this domain during gating. □

Methods

Mutagenesis

Point mutations were created in the complementary DNA encoding the human GABA_A-R α_1 subunit using the QuikChange kit (Stratagene) as described²⁵ and sequences of the mutated cDNA inserts were confirmed by automated fluorescent DNA sequencing (Cornell University, Ithaca, NY, and Rockefeller University, New York, NY); mutagenic primer sequences are available from N.L.H. on request. GABA_A receptors were expressed in HEK 293 cells, maintained and transfected as described²⁵.

Electrophysiology

Recordings were made 48–72 h after transfection at room temperature (20–22 °C), using the whole-cell patch-clamp technique²⁵. Concentration–response curves were determined for GABA or P4S in wild-type and mutant GABA_A-Rs as described¹³. We defined the relative efficacy (ϵ) of P4S by $\epsilon = I_{\max}(\text{P4S})/I_{\max}(\text{GABA})$, where $I_{\max}(\text{P4S})$ is the maximum current elicited by a saturating concentration of P4S, and $I_{\max}(\text{GABA})$ the current elicited by a saturating concentration of GABA. Crosslinking experiments were done as described²¹. All data are reported as the mean $\pm$ s.e.m. Statistical significance was determined using analysis of variance or a Student's two-tailed, unpaired *t*-test. A $\Delta\Delta G$ value was calculated for each mutant receptor using the formula: $\Delta\Delta G = RT \ln(\text{mutant } EC_{50}/\text{wild-type } EC_{50})$, and the coupling coefficient (Ω) for each double mutant was calculated as described¹⁸. For ease of comparison, we have reported the value of Ω^{-1} when $\Omega < 1$.

Molecular modelling

We built a model of the extracellular domain of a GABA_A-R by threading the sequence of the GABA_A-R α_1 subunit (GABRA1) onto the crystal structure of the acetylcholine binding protein (AChBP) as described²⁶. In brief, coordinates of the AChBP⁴ were obtained from the Protein Data Bank (accession code 1I9B). This structure contains 205 amino acid residues that have 24% sequence identity with nAChR $\alpha 7$ and 15–18% identity with GABA_A-R α -subunits. The GABRA1 sequence was aligned with the AChBP template⁴, and the backbone atoms of the sequence were assigned the corresponding coordinates from the crystal structure of the AChBP using the Homology Module of Insight 2000 (Accelrys). We built loops to fill in the gaps between the GABRA1 sequence and the template sequence and refined the model using the auto-rotamer feature in the Biopolymer Module of Insight 2000. The backbone atoms (C, C α , N) of each GABRA1 residue were tethered to the coordinates of corresponding residues in the AChBP template

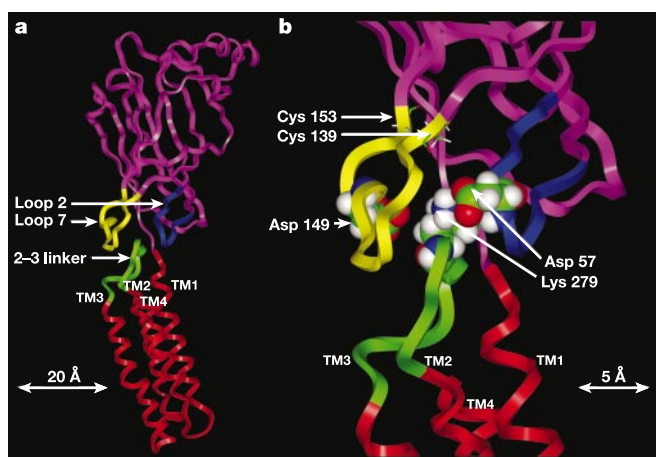


Figure 3 Molecular model of the GABA_A-R α_1 subunit. **a**, Model of a whole GABA_A-R α_1 subunit, viewed from the centre of the ion channel. The extracellular loop 2 is highlighted in blue, loop 7 in yellow, and the 2–3L region in green. The rest of the extracellular domain is shown in magenta and the transmembrane domain, modelled here as a bundle of four α -helices, is shown in red. Scale bar, 20 Å. **b**, Detailed view of the potential contacts between the extracellular and the transmembrane domains. Extracellular β -strands and transmembrane α -helices are shown as magenta and red ribbons, respectively. Several key residues have been rendered with a space-filling surface for easy identification: The basic 2–3L residue K279 and the acidic residues in loop 2 (D57) and loop 7 (D149). Residues C139 and C153 define loop 7 (the signature loop). Scale bar, 5 Å.

with a force constant of $5 \text{ kcal } \text{\AA}^{-2}$ and the structure optimized with the Discover 2000 module of Insight 2000 (ref. 26).

We built a model of the whole GABA_A-R α -subunit by superimposing this model of the extracellular domain onto a model of the transmembrane domain of GABRA1 (ref. 23). Briefly, the two models were superimposed along their pentameric axes of symmetry and moved vertically toward each other until they were in van der Waals contact. The model of the ligand-binding domain was rotated about its axis of symmetry until the two cysteine residues that define the cysteine loop (Cys 139 and Cys 153) were directly over the α -helix representing TM3 in the model of the transmembrane domain as shown in a cryoelectron microscopy image of the nAChR²³. The two models were then merged and the structure was optimized with Discover 2000 (ref. 27).

Received 7 August; accepted 21 October 2002; doi:10.1038/nature01280.

- Schofield, P. R. *et al.* Sequence and functional expression of the GABA_A receptor shows a ligand-gated receptor super-family. *Nature* **328**, 221–227 (1987).
- Betz, H. Ligand-gated ion channels in the brain: the amino acid receptor superfamily. *Neuron* **5**, 383–392 (1990).
- Karlin, A. Emerging structure of the nicotinic acetylcholine receptors. *Nature Rev. Neurosci.* **3**, 102–114 (2002).
- Brejci, K. *et al.* Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **411**, 269–276 (2001).
- Smith, G. B. & Olsen, R. W. Identification of a [³H]muscimol photoaffinity substrate in the bovine γ -aminobutyric acid_A receptor α subunit. *J. Biol. Chem.* **269**, 20380–20387 (1994).
- Sigel, E., Baur, R., Kellenberger, S. & Malherbe, P. Point mutations affecting antagonist affinity and agonist dependent gating of GABA_A receptor channels. *EMBO J.* **11**, 2017–2023 (1992).
- Boileau, A. J., Evers, A. R., Davis, A. F. & Czajkowski, C. Mapping the agonist binding site of the GABA_A receptor: evidence for a β -strand. *J. Neurosci.* **19**, 4847–4854 (1999).
- Grosman, C., Zhou, M. & Auerbach, A. Mapping the conformational wave of acetylcholine receptor channel gating. *Nature* **403**, 773–776 (2000).
- Dougherty, D. A. & Lester, H. A. Neurobiology. Snails, synapses and smokers. *Nature* **411**, 252–253 (2001).
- Croxen, R. C. *et al.* Mutations in different functional domains of the human muscle acetylcholine receptor α subunit in patients with the slow-channel congenital myasthenic syndrome. *Hum. Mol. Genet.* **6**, 767–774 (1997).
- Shiang, R. *et al.* Mutations in the $\alpha 1$ subunit of the inhibitory glycine receptor cause the dominant neurologic disorder, hyperekplexia. *Nature Genet.* **4**, 351–358 (1993).
- Baulac, S. *et al.* First genetic evidence of GABA_A receptor dysfunction in epilepsy: mutation in the $\gamma 2$ -subunit gene. *Nature Genet.* **28**, 46–48 (2001).
- O'Shea, S. M. & Harrison, N. L. Arg-274 and Leu-277 of the γ -aminobutyric acid type A receptor $\alpha 2$ subunit define agonist efficacy and potency. *J. Biol. Chem.* **275**, 22764–22768 (2000).
- Sigel, E., Buhr, A. & Baur, R. Role of the conserved lysine residue in the middle of the predicted extra-cellular loop between M2 and M3 in the GABA_A receptor. *J. Neurochem.* **73**, 1758–1764 (1999).
- Ebert, B. *et al.* Differences in agonist/antagonist binding affinity and receptor transduction using recombinant human γ -aminobutyric acid type A receptors. *Mol. Pharmacol.* **52**, 1150–1156 (1997).
- Steinbach, J. H. & Akk, G. Modulation of GABA_A receptor channel gating by pentobarbital. *J. Physiol. (Lond.)* **537**, 715–733 (2001).
- Carter, P. J., Winter, G., Wilkinson, A. J. & Fersht, A. R. The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Cell* **38**, 835–840 (1984).
- Hidalgo, P. & MacKinnon, R. Revealing the architecture of a K⁺ channel pore through mutant cycles with a peptide inhibitor. *Science* **268**, 307–310 (1995).
- Schreiber, G. & Fersht, A. R. Energetics of protein–protein interactions: analysis of the barnase–barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* **248**, 478–486 (1995).
- DiPaola, M., Czajkowski, C. & Karlin, A. The sidedness of the COOH terminus of the acetylcholine receptor δ subunit. *J. Biol. Chem.* **264**, 15457–15463 (1989).
- Horenstein, J., Wagner, D. A., Czajkowski, C. & Akabas, M. H. Protein mobility and GABA-induced conformational changes in GABA_A receptor pore-lining M2 segment. *Nature Neurosci.* **4**, 477–485 (2001).
- Thornton, J. M. Disulphide bridges in globular proteins. *J. Mol. Biol.* **151**, 261–287 (1981).
- Bertaccini, E. & Trudell, J. R. Predicting the transmembrane secondary structure of ligand-gated ion channels. *Protein Eng.* **15**, 443–454 (2002).
- Unwin, N., Miyazawa, A., Li, J. & Fujiyoshi, Y. Activation of the nicotinic acetylcholine receptor involves a switch in conformation of the α subunits. *J. Mol. Biol.* **319**, 1165–1176 (2002).
- Nishikawa, K., Jenkins, A., Paraskevaki, I. & Harrison, N. L. Volatile anesthetic actions on the GABA_A receptors: contrasting effects of $\alpha 1$ (S270) and $\beta 2$ (N265) point mutations. *Neuropharmacology* **42**, 337–345 (2002).
- Trudell, J. Unique assignment of inter-subunit association in GABA_A $\alpha 1\beta 5\gamma 2$ receptors determined by molecular modelling. *Biochim. Biophys. Acta* **1565**, 91–96 (2002).
- Trudell, J. R. & Bertaccini, E. Molecular modelling of specific and non-specific anaesthetic interactions. *Br. J. Anaesth.* **89**, 32–40 (2002).
- Bera, M., Chatav, M. & Akabas, M. GABA(A) receptor M2-3 loop secondary structure and changes in accessibility during channel gating. *J. Biol. Chem.* **277**, 43002–43010 (2002).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank J. Horenstein for advice on crosslinking experiments and the NIH for financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.L.H. (e-mail: neh2001@med.cornell.edu).

Periodic Notch inhibition by Lunatic Fringe underlies the chick segmentation clock

J. K. Dale*†, M. Maroto*†, M.-L. Dequeant†, P. Malapert†, M. McGrew† & O. Pourquie†

Laboratoire de génétique et de physiologie du développement, Institut de biologie du développement de Marseille, CNRS-INSERM-Université de la méditerranée-AP de Marseille, Campus de Luminy, Case 907, 13288 Marseille Cedex 09, France

* These authors contributed equally to this work

† Present addresses: Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, Missouri 64110, USA (J.K.D., M.M., M.-L.D., P.M. and O.P.); Department of Gene Expression and Development, Roslin Institute, Roslin EH25 9PS, UK (M.McG.)

The segmented aspect of the vertebrate body plan first arises through the sequential formation of somites. The periodicity of somitogenesis is thought to be regulated by a molecular oscillator, the segmentation clock, which functions in presomitic mesoderm cells. This oscillator controls the periodic expression of 'cyclic genes', which are all related to the Notch pathway^{1–7}. The mechanism underlying this oscillator is not understood. Here we show that the protein product of the cyclic gene *lunatic fringe* (*Lfng*), which encodes a glycosyltransferase that can modify Notch activity, oscillates in the chick presomitic mesoderm. Overexpressing *Lfng* in the paraxial mesoderm abolishes the expression of cyclic genes including endogenous *Lfng* and leads to defects in segmentation. This effect on cyclic genes phenocopies inhibition of Notch signalling in the presomitic mesoderm. We therefore propose that *Lfng* establishes a negative feedback loop that implements periodic inhibition of Notch, which in turn controls the rhythmic expression of cyclic genes in the chick presomitic mesoderm. This feedback loop provides a molecular basis for the oscillator underlying the avian segmentation clock.

Cyclic genes include downstream targets of Notch such as transcription factors of the *hairy/E(spl)* family^{4–6}, the glycosyltransferase *Lfng*^{1–3} and the Notch ligand *Delta*^{C7}. Cyclic gene messenger RNA expression appears as a wave that sweeps the presomitic mesoderm (PSM) once during the formation of each somite^{8,9}. The correlation between the waves of mRNA expression across the PSM and somitogenesis itself suggests that the segmentation clock has a causal role in somite formation, although direct evidence for this is lacking. Because the segmentation clock is responsible for the rhythmic expression of Notch targets, one role of the clock is thought to be a periodic modulation of Notch activity involved in generating the somitic boundaries⁸. Accordingly, in zebrafish, mice and chick, several members of the Notch pathway, including receptors and ligands, are expressed broadly in the PSM, and the phenotypes of mice and fish embryos carrying null mutations in different members of this pathway show somitic defects^{7,10–12}. In regard to Notch function as part of the oscillation mechanism itself, conflicting data in zebrafish indicate that Notch signalling has a role in either regulating the mRNA oscillations¹⁰ or their synchronization⁷, whereas analyses of mice with mutations in genes such as *Delta-like 1* or *RBPjK* show a marked downregulation and an apparent loss of the dynamic aspect of cyclic gene expression in the PSM^{5,13,14}. Thus, it remains unclear whether Notch signalling acts as a cofactor for cyclic gene expression (part of the clock output), or as part of the core mechanism of the segmentation clock (a clock component).

The expression sequence of the cyclic genes has been divided into

three different spatial patterns that follow one another cyclically. In phase I, the expression domain is broad and extends through the caudal half of the PSM; in phase II, the domain shifts towards the middle of the PSM; and in phase III, the domain becomes a narrow band in the rostral-most PSM^{4,9}. To determine whether the expression sequence of *Lfng* traverses the chick PSM as a continuous progressive wave or as discrete blocks, we analysed the expression sequence of *Lfng* mRNA precisely. Several embryos ($n = 89$) containing 14–18 somites were hybridized with a *Lfng* riboprobe, and measurements of the domains of *Lfng* expression in the PSM and their distance from the last somite boundary were scored. Because the wave of expression moves anteriorly⁴, the embryos were aligned in order of the advancing position of the anterior front of the expression sequence (Fig. 1a). The data show that *Lfng* mRNA expression does indeed sweep across the PSM in a progressive manner. From the number of embryos observed in each phase, we estimate that expression in the caudal PSM, phase I domain (22 of 89 embryos), lasts for about 20–25% of the cycle, whereas phases II and III make up the remaining 75–80% of the cycle.

To examine whether the protein products of the cyclic genes are also expressed periodically in the PSM, we generated polyclonal antibodies against a polypeptide derived from the sequence of chick *Lfng*. These antibodies recognize a unique band of the expected size in western blot analysis of either chick PSM cells or cells transfected with a chick *Lfng* construct (data not shown). We subsequently analysed the dynamics of *Lfng* expression in the PSM by comparing the protein profile in the posterior half of one PSM (by western blot) with the RNA expression profile in the contralateral PSM of the

same embryo (by *in situ* hybridization; $n = 35$). We restricted the analysis to the posterior PSM because, unlike the anterior PSM, cells show on–off oscillations of the cyclic genes throughout this domain. This analysis showed that the quantities of *Lfng* vary in the posterior half of the PSM as compared with β -actin (Fig. 1b). Thus, *Lfng* protein is at a minimum in early phase II and rises to a maximum in early phase III (Fig. 1c). Such a rapid protein turnover is consistent with *Lfng* functioning as a clock component with a role in the core mechanism of the oscillator. Similar oscillations with a 2-h periodicity have been shown for *Hes1* in serum-shocked mouse cells¹⁵.

To test whether this dynamic expression of *Lfng* protein is required for the clock mechanism, we misexpressed *Lfng* in the chick PSM using *in ovo* electroporation and analysed the effect on cyclic genes¹⁶. The expression profile of endogenous *Lfng* mRNA was analysed using a probe that does not recognize the *Lfng* transcript from the electroporated construct. Of 49 electroporated controls, 14 embryos (28%) were observed to express *Lfng*, *c-hairy1* or *c-hairy2* in a phase I domain (Fig. 2a and data not shown). This ratio was similar to that observed in several non-electroporated embryos (24.7%; Student's *t*-test, $P = 0.6249$; Fig. 1a). By contrast, of 19 embryos electroporated with pCAGGS–*Lfng*, none showed expression of *Lfng*, *c-hairy1* or *c-hairy2* mRNA in phase I (Fig. 2b and data not shown). But expression of these genes in other tissues (see Fig. 2) and in areas outside the green fluorescent protein (GFP)-positive domain was unchanged as compared with controls (data not shown). *Tbx6L* expression was not affected by misexpression of *Lfng*, indicating that PSM identity was maintained ($n = 3$; data not shown). Thus, our observations suggest that misexpression of *Lfng* abolishes cyclic gene expression in the caudal PSM, including expression from the endogenous *Lfng* locus ($P < 0.01$). Consequently, this implies that *Lfng* negatively regulates its own mRNA expression, potentially providing a molecular basis for a feedback loop underlying the segmentation clock mechanism.

We examined the effect of *Lfng* overexpression on somitogenesis, by developing some of the electroporated embryos until the region of the PSM in which cycling gene expression was disrupted had

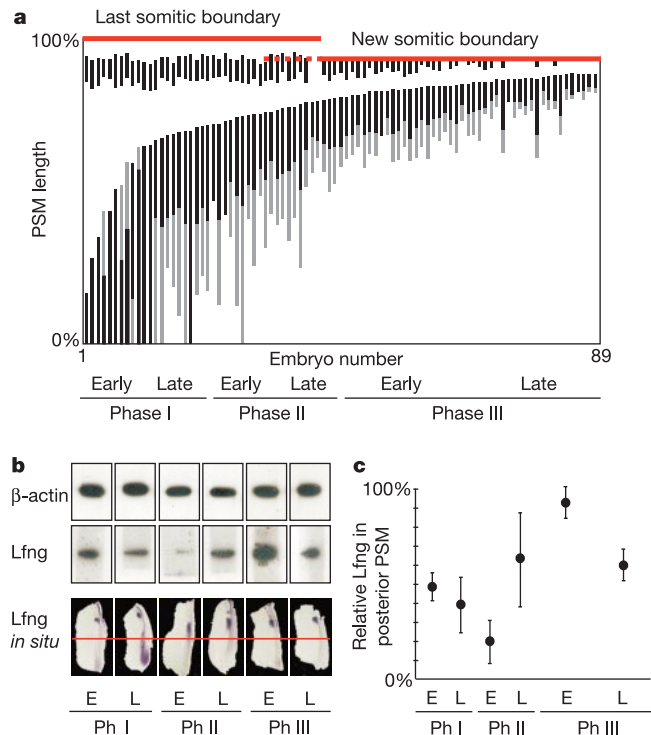


Figure 1 *Lfng* protein oscillates in the chick PSM. **a**, Each bar represents one embryo (x axis). The expression domains along the A–P axis of the PSM are represented as bars plotted along the y axis, anterior to the top. Darker bars represent stronger expression. **b**, Representative embryos at different phases of the cycle. The top two rows show the amount of β -actin and *Lfng* protein in the same posterior PSM. The bottom row represents the *Lfng* mRNA expression profile in the contralateral side of the corresponding embryo. **c**, Graphical representation of the relative changes in *Lfng* protein in the posterior PSM (y axis) during the *Lfng* mRNA expression cycle (x axis).

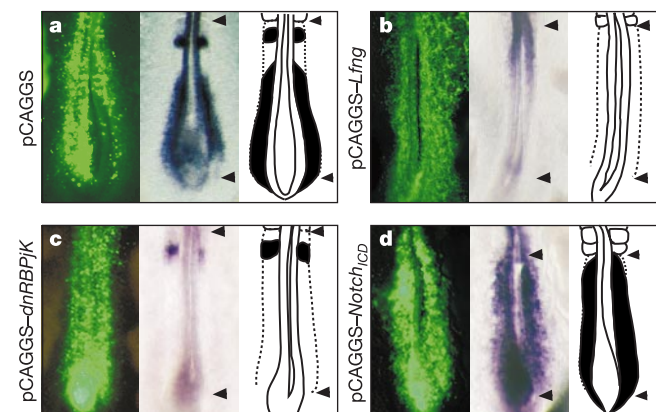


Figure 2 Misexpression of *Lfng* or *dnRBPJk* abolishes cyclic expression of *Lfng*, whereas *Notch1cd* causes ectopic expression of *Lfng*. Embryos were electroporated with pCAGGS–GFP plus pCAGGS (**a**), pCAGGS–*Lfng* (**b**), pCAGGS–*dnRBPJk* (**c**) or pCAGGS–*Notch1cd* (**d**) and hybridized for *Lfng* expression. Left, GFP in the PSM and somites. Right, expression of *Lfng* in the PSM (hatched area). Arrows mark the rostral and caudal PSM limits. Note that, unlike embryos misexpressing *Lfng*, many of those misexpressing *dnRBPJk* maintain a band of *Lfng* expression in the anterior PSM. This may be indicative of a difference in the regulation and the function of the clock genes in this domain^{14,27}.

segmented. *Lfng* misexpression severely perturbed the positioning of somite boundaries ($n = 22$; compare the central panels in Fig. 3a and d), a phenotype similar to that of *lunatic fringe* null mice^{17,18}. These results suggest that the rhythmic oscillations of Lunatic Fringe protein, and not the presence of the protein *per se*, are essential for coordinating boundary positioning. Expression of *Uncx4.1* and *c-Delta1*, which are markers of the posterior somitic compartment (refs 11, 19, and Fig. 3a and b), was disorganized and mosaic throughout the somites in embryos overexpressing *Lfng* ($n = 18$; Fig. 3d and e). By contrast, dermomyotome and sclerotome formation occurred normally (Fig. 3d), and the onset of *MyoD* expression was normal ($n = 4$; Fig. 3f). These data therefore indicate that disruption of cyclic gene expression, mediated by misexpression of *Lfng*, is associated with the irregular positioning of somitic boundaries and a disruption of antero-posterior somitic compartmentalization, although somite differentiation seems to be largely unaffected. Because *Lfng* has been implicated in somite boundary formation²⁰, however, we cannot rule out the possibility that these segmentation defects (Fig. 3) may result in part from a later interference with this process rather than from a disruption of the segmentation clock function.

The Fringe proteins are glycosyltransferases, which modify Notch–ligand interactions to inhibit or potentiate Notch signalling^{21–24}. In mouse Notch signalling mutants, a downregulation of *Lfng* mRNA expression is seen in the PSM^{13,14}. Thus, the downregulation of cyclic gene expression that we observed indicated that *Lfng* might inhibit Notch signalling in the chick PSM. To investigate this issue, we compared the phenotypic effect of *Lfng* misexpression to that of disrupting the Notch pathway.

We analysed the effect on *Lfng* mRNA expression in the PSM after driving expression of a dominant-negative form of RBPjk in this tissue by *in ovo* electroporation²⁵. Of 16 electroporated embryos, 14 showed no caudal PSM expression and 2 showed severely down-regulated caudal expression of *Lfng* (Fig. 2c). By contrast, *Tbx6L* expression was not affected ($n = 2$; data not shown). These data show that blocking RBPJK-dependent Notch signalling inhibits the mRNA expression of *Lfng* in the caudal PSM ($P < 0.05$), similar to the effect seen when misexpressing *Lfng*.

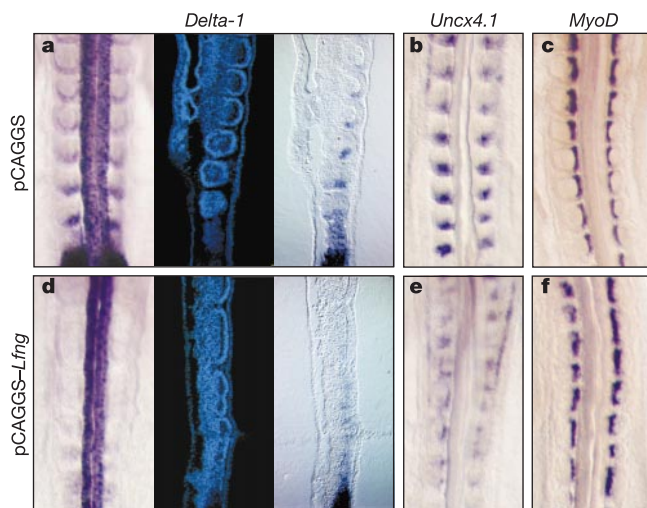


Figure 3 *Lfng* misexpression severely disrupts positioning of somitic borders and antero-posterior somitic compartmentalization. Dorsal view is shown of embryos electroporated with pCAGGS and pCAGGS–GFP (a–c), or pCAGGS–*Lfng* and pCAGGS–GFP (d–f) and hybridized for mRNA expression of *c-Delta1* (a, d), *Uncx4.1* (b, e) or *MyoD* (c, f). Included in a and c are sagittal sections of the same whole-mount embryo to visualize Hoechst staining (middle) or expression of *c-Delta1* mRNA (right).

We then drove expression of a constitutively active form of Notch (Notch_{ICD})²⁶ throughout the PSM. Although *Tbx6L* expression was unchanged ($n = 2$; data not shown), 8 of 12 embryos (67%) showed a strong overexpression of *Lfng* along the whole PSM, including the phase I domain (Fig. 2d). Thus, constitutive activation of Notch signalling can drive ectopic expression of *Lfng* ($P < 0.05$), indicating that *Lfng* is a Notch target in the chick PSM, as has been shown in mice^{14,27}. Together, these data indicate that cyclic Notch activation may have an essential role in generating oscillations of *Lfng* mRNA in the chick PSM.

We examined the requirement for γ -secretase-mediated cleavage of the Notch receptor in driving *Lfng* mRNA oscillations. Half embryos were cultured for the same time period with one side acting as a control, while the other was treated with γ -secretase inhibitors such as MW167, DAPT and MG132 (ref. 28; and M. Wolfe, personal communication). After culture, no reactivation of the phase I domain of *Lfng* mRNA expression was seen as compared with the control side ($n = 30$; Fig. 4a). These data indicate that preventing the release of the intracytoplasmic domain of Notch blocks, rather than desynchronizes, the oscillating mRNA expression of *Lfng*. By contrast, expression of *Tbx6L* and *paraxis* was maintained ($n = 11$; Fig. 4b and data not shown). In addition, *Lfng* expression in the caudal PSM could be recovered by re-incubating the explants after the reagent had been washed away ($n = 10$; Fig. 4c). Specificity of the drugs effect on Notch signalling was verified by misexpressing Notch_{ICD} in the PSM by electroporation before drug treatment ($n = 11$; Fig. 4d). Under these conditions, *Lfng* mRNA expression was rescued in the caudal PSM of the inhibitor-treated explant. Together, these results show that blocking the release of the Notch intracytoplasmic domain phenocopies the effect seen by misexpressing *Lfng* in the PSM.

In the vertebrate embryo, the only Notch ligands expressed along the PSM belong to the Delta family. Thus, our observations indicating that *Lfng* mediates inhibition of Notch signalling in the PSM are at odds with the situation in the fly imaginal disks, in which

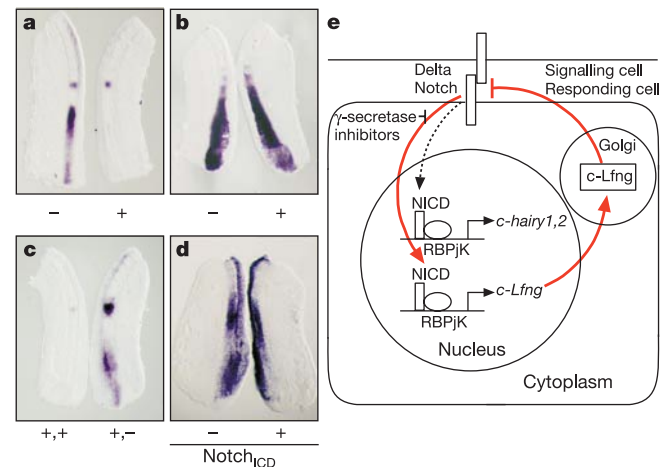


Figure 4 Blocking Notch cleavage blocks the segmentation clock. a, b, Expression of *Lfng* (a) or *Tbx6L* (b) in explants cultured in the presence (+) or absence (–) of MW167 or DAPT, respectively. c, *Lfng* expression in explants cultured with MG132 and then further cultured in the presence (+, +) of MG132 (left explant) or washed and cultured in its absence (+, –) (right explant). d, *Lfng* expression in explants dissected from an embryo misexpressing Notch_{ICD}, and then cultured with or without DAPT. e, Model of an oscillator mechanism based on negative feedback. Notch signalling activates cyclic gene transcription. *Lfng* then closes the loop by modifying Notch, thereby inhibiting Notch signalling. This negative effect is transient owing to the rapid turnover of *Lfng*. NICD, Notch_{ICD}.

Fringe potentiates Delta-mediated Notch activation²¹. Our data identify a negative feedback loop acting in the chick PSM in which Notch-dependent production of the *Lfng* protein will act, in turn, to inhibit Notch signalling and thus downregulate *Lfng* expression. We have shown that the protein product of *Lfng* oscillates in presomitic cells with the same periodicity as somitogenesis. Thus, because the *Lfng* protein has a rapid turnover, its negative effect on Notch signalling will be transient and periodic (Fig. 4e). We propose that this negative feedback loop represents a core component of the avian segmentation clock mechanism. □

Methods

Plasmids and *in situ* hybridization

Whole-mount *in situ* hybridization was done as described²⁹. The 3' untranslated region *Lfng* probe was produced by polymerase chain reaction (PCR) amplification of a fragment corresponding to nucleotides 1,719–2,567 of the chick *Lfng* cDNA sequence (GenBank accession code U91849). The 400-base-pair chicken *Uncc4.1* probe was produced by PCR amplification from the ChEST478F8 clone, found in the BursalEst Database maintained by J.-M. Buerstedde, Heinrich-Pette Institute for Experimental Virology and Immunology, Hamburg, Germany. We prepared plasmids for *in ovo* electroporation as described¹⁶.

In ovo electroporation

In ovo electroporation was done as described¹⁶. We used only those embryos showing both a normal morphology outside the GFP-positive domain and a high intensity of GFP expression in the PSM analysis by *in situ* hybridization. Thus, between 182 and 631 embryos were electroporated with each construct, and of these 5–13% were analysed. The statistical significance of the differences in the number of embryos showing a phase I domain of expression when comparing non-electroporated versus control electroporated embryos, or control electroporated versus experimental electroporated constructs, was calculated using Student's *t*-test at a 95–99% confidence level.

Graphical analysis

Measurements of the different domains of *Lfng* mRNA expression in the PSM and their distance from the last-formed somite boundary were scored. These values were then transformed into a graphical representation of the PSM domain of all 89 embryos analysed, using Excel (Microsoft). We interpreted the percentage of embryos in each phase as the percentage duration of each phase during the 90-min cycle.

Antibody production

A PCR fragment corresponding to nucleotides 493–1,092 of the chick *Lfng* ORF was cloned into the expression vector pET-23b (Novagen) using DH5 α *Escherichia coli* as a bacterial host. The BL21(DE3) *E. coli* strain was transformed with the construct pET-23b-cLfng and protein production was induced by addition of 0.5 mM IPTG. Recombinant protein was purified as inclusion bodies that were used to prepare rabbit polyclonal antibodies (Eurogentec).

Analysis of *Lfng* protein expression profile

We treated each embryo independently. One side was immediately fixed and analysed by *in situ* hybridization using a *Lfng* probe to compare *Lfng* mRNA and protein profiles in the same embryo. Only the posterior half of the PSM from the contralateral side of the same embryo was dissected in cold PBS containing the Complete cocktail of protease inhibitors (Boehringer Mannheim). We carried out western blot analysis on this sample using a monoclonal antibody against β -actin (A5441; Sigma) or an antibody against c-Lfng and the membranes were visualized using electrochemiluminescence. We scanned and quantified films using NIH image software. After normalizing the total amount of protein in each sample by comparison to β -actin, the corresponding amounts of *Lfng* protein in individual samples were compared within each experiment as a ratio with respect to the sample showing the maximum *Lfng* value. These values were then grouped according to mRNA expression in the control side. The mean value and standard deviation were calculated and plotted as a graphical representation of the fluctuating *Lfng* quantities during the 90-min cycle (Fig. 1d). Note that the antibody does not work in immunocytochemical assessments of either whole-mount or sections of chick embryos.

γ -Secretase inhibitor treatment and half-embryo assay

We used MW167 (ref. 28; Enzyme Systems), DAPT (a gift from M. Wolfe) and MG132 (ref. 28; Calbiochem) to block the cleavage of Notch. These reagents were diluted into culture medium from stock solutions in dimethylsulphoxide (DMSO) to maintain a final DMSO concentration of 1%. Half embryo assays were done as described⁴. One half was cultured in medium containing MW167 (20 μ M), DAPT (10 μ M) or MG132 (100 μ M), whereas the control side was cultured in normal medium. Both sides were cultured for 2–4 h and then analysed for expression of *Lfng* mRNA by *in situ* hybridization. Reversibility of the drugs was shown by culturing explants for 2 h 15 min in the presence of MG132 (100 μ M), followed by culturing one explant for a further 1 h 20 min in the presence of MG132, while washing and then culturing the second explant for 1 h 20 min in the absence of MG132.

Received 22 July; accepted 21 October 2002; doi:10.1038/nature01244.

Published online 12 January 2003.

- McGrew, M. J., Dale, J. K., Fraboulet, S. & Pourquie, O. The *lunatic fringe* gene is a target of the molecular clock linked to somite segmentation in avian embryos. *Curr. Biol.* **8**, 979–982 (1998).
- Aulehla, A. & Johnson, R. L. Dynamic expression of *lunatic fringe* suggests a link between notch signaling and an autonomous cellular oscillator driving somite segmentation. *Dev. Biol.* **207**, 49–61 (1999).
- Forsberg, H., Crozet, F. & Brown, N. A. Waves of mouse *Lunatic fringe* expression, in four-hour cycles at two-hour intervals, precede somite boundary formation. *Curr. Biol.* **8**, 1027–1030 (1998).
- Palmeirim, I., Henrique, D., Ish-Horowicz, D. & Pourquie, O. Avian *hairly* gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell* **91**, 639–648 (1997).
- Jouve, C. *et al.* Notch signalling is required for cyclic expression of the *hairly*-like gene *HES1* in the presomitic mesoderm. *Development* **127**, 1421–1429 (2000).
- Holley, S. A., Geisler, R. & Nusslein-Volhard, C. Control of *her1* expression during zebrafish somitogenesis by a delta-dependent oscillator and an independent wave-front activity. *Genes Dev.* **14**, 1678–1690 (2000).
- Jiang, Y. J. *et al.* Notch signalling and the synchronization of the somite segmentation clock. *Nature* **408**, 475–479 (2000).
- Maroto, M. & Pourquie, O. A molecular clock involved in somite segmentation. *Curr. Top. Dev. Biol.* **51**, 221–248 (2001).
- Pourquie, O. & Tam, P. P. A nomenclature for prospective somites and phases of cyclic gene expression in the presomitic mesoderm. *Dev. Cell* **1**, 619–620 (2001).
- Holley, S. A., Julich, D., Rauch, G. J., Geisler, R. & Nusslein-Volhard, C. *her1* and the *notch* pathway function within the oscillator mechanism that regulates zebrafish somitogenesis. *Development* **129**, 1175–1183 (2002).
- Brabe de Angelis, M., McIntyre, J. & Gossler, A. Maintenance of somite borders in mice requires the Delta homologue Dll1. *Nature* **386**, 717–721 (1997).
- Conlon, R. A., Reaume, A. G. & Rossant, J. Notch1 is required for the coordinate segmentation of somites. *Development* **121**, 1533–1545 (1995).
- del Barco Barrantes, I. *et al.* Interaction between Notch signalling and Lunatic Fringe during somite boundary formation in the mouse. *Curr. Biol.* **9**, 470–480 (1999).
- Morales, A. V., Yasuda, Y. & Ish-Horowicz, D. Periodic Lunatic fringe expression is controlled during segmentation by a cyclic transcriptional enhancer responsive to Notch signaling. *Dev. Cell* **3**, 63–74 (2002).
- Hirata, H. *et al.* Oscillatory expression of the bHLH factor *Hes1* regulated by a negative feedback loop. *Science* **298**, 840–843 (2002).
- Dubrulle, J., McGrew, M. J. & Pourquie, O. FGF signaling controls somite boundary position and regulates segmentation clock control of spatiotemporal *Hox* gene activation. *Cell* **106**, 219–232 (2001).
- Evrard, Y. A., Lun, Y., Aulehla, A., Gan, L. & Johnson, R. L. *lunatic fringe* is an essential mediator of somite segmentation and patterning. *Nature* **394**, 377–381 (1998).
- Zhang, N. & Gridley, T. Defects in somite formation in *lunatic fringe*-deficient mice. *Nature* **394**, 374–377 (1998).
- Mansouri, A. *et al.* Paired-related murine homeobox gene expressed in the developing sclerotome, kidney, and nervous system. *Dev. Dyn.* **210**, 53–65 (1997).
- Sato, Y., Yasuda, K. & Takahashi, Y. Morphological boundary forms by a novel inductive event mediated by Lunatic fringe and Notch during somitic segmentation. *Development* **129**, 3633–3644 (2002).
- Panin, V. M., Papayannopoulos, V., Wilson, R. & Irvine, K. D. Fringe modulates Notch–ligand interactions. *Nature* **387**, 908–912 (1997).
- Munro, S. & Freeman, M. The notch signalling regulator fringe acts in the Golgi apparatus and requires the glycosyltransferase signature motif DXD. *Curr. Biol.* **10**, 813–820 (2000).
- Moloney, D. J. *et al.* Fringe is a glycosyltransferase that modifies Notch. *Nature* **406**, 369–375 (2000).
- Bruckner, K., Perez, L., Clausen, H. & Cohen, S. Glycosyltransferase activity of Fringe modulates Notch–Delta interactions. *Nature* **406**, 411–415 (2000).
- Waltzer, L., Bourillot, P., Sergeant, A. & Manet, E. RBP-jk repression activity is mediated by a co-repressor and antagonised by the Epstein Barr virus transcription factor EBNA2. *Nucleic Acids Res.* **23**, 4939–4945 (1995).
- Schroeter, E. H., Kisslinger, J. A. & Kopan, R. Notch-1 signalling requires ligand-induced proteolytic release of intracellular domain. *Nature* **393**, 382–386 (1998).
- Cole, S. E., Levorse, J. M., Tilghman, S. M. & Vogt, T. F. Clock regulatory elements control cyclic expression of Lunatic Fringe during somitogenesis. *Dev. Cell* **3**, 75–84 (2002).
- De Strooper, B. *et al.* A presenilin-1-dependent γ -secretase-like protease mediates release of Notch intracellular domain. *Nature* **398**, 518–522 (1999).
- Henrique, D. *et al.* Expression of a Delta homologue in prospective neurons in the chick. *Nature* **375**, 787–790 (1995).

Acknowledgements We thank A.-C. Petit for help with analysing the *Lfng* mRNA expression profile; J. P. Rey for technical assistance; K. Katsube, D. Henrique, S. Mackem, O. Saitoh, E. Olson, M. Wolfe, A. Israel and R. Kopan for reagents; P. Torregrossa for help with statistical analysis; R. Kageyama for sharing unpublished information; T. Lecuit, S. Kerridge, D. Ish-Horowicz, M. Wolfe, M. Placzek, C. Hudson, J. Dubrulle, T. Imura, S. Millet and V. Baubet for comments on the manuscript. This work was supported by a Wellcome Prize Travelling Research Fellowship, an EMBO Long term Fellowship, and a Fondation pour la recherche médicale (FRM) Postdoctoral Fellowship to J.K.D., a Marie Curie Individual Fellowship from the European Commission to M.M., and the laboratory was supported by funding from Centre national de la recherche scientifique (CNRS), Human Frontiers Science Programme Organisation (HESPO), Association française contre les myopathies (AFM) and the Université de la méditerranée-AP de Marseille.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to O.P. (e-mail: olp@stowers-institute.org).

A zebrafish homologue of the chemokine receptor *Cxcr4* is a germ-cell guidance receptor

Holger Knaut*, Christian Werz*, Robert Geisler*, The Tübingen 2000 Screen Consortium† & Christiane Nüsslein-Volhard*

* Max Planck Institut für Entwicklungsbiologie, Abteilung III/Genetik, Spemannstrasse 35, 72076 Tübingen, Germany

† A complete list of authors appears at the end of the paper

Germ cells preserve an individual's genetic information and transmit it to the next generation. Early in development germ cells are set aside and undergo a specialized developmental programme, a hallmark of which is the migration from their site of origin to the future gonad¹. In *Drosophila*, several factors have been identified that control germ-cell migration to their

target tissues^{2–4}; however, the germ-cell chemoattractant or its receptor have remained unknown. Here we apply genetics and *in vivo* imaging to show that *odysseus*, a zebrafish homologue of the G-protein-coupled chemokine receptor *Cxcr4*, is required specifically in germ cells for their chemotaxis. *odysseus* mutant germ cells are able to activate the migratory programme, but fail to undergo directed migration towards their target tissue, resulting in randomly dispersed germ cells. SDF-1, the presumptive cognate ligand for *Cxcr4*, shows a similar loss-of-function phenotype and can recruit germ cells to ectopic sites in the embryo, thus identifying a vertebrate ligand–receptor pair guiding migratory germ cells at all stages of migration towards their target.

In search of factors that affect the specification, maintenance or migration of primordial germ cells (PGCs) in zebrafish, we assayed embryos from a large-scale genetic screen for PGC number and position using *vasa*, a PGC-specific marker^{5–7}. In zebrafish, PGCs originate at random positions with respect to the body axis. At 14 h after fertilization they align with the presomitic mesoderm resulting in one cluster on either side of the embryo. The PGCs then migrate

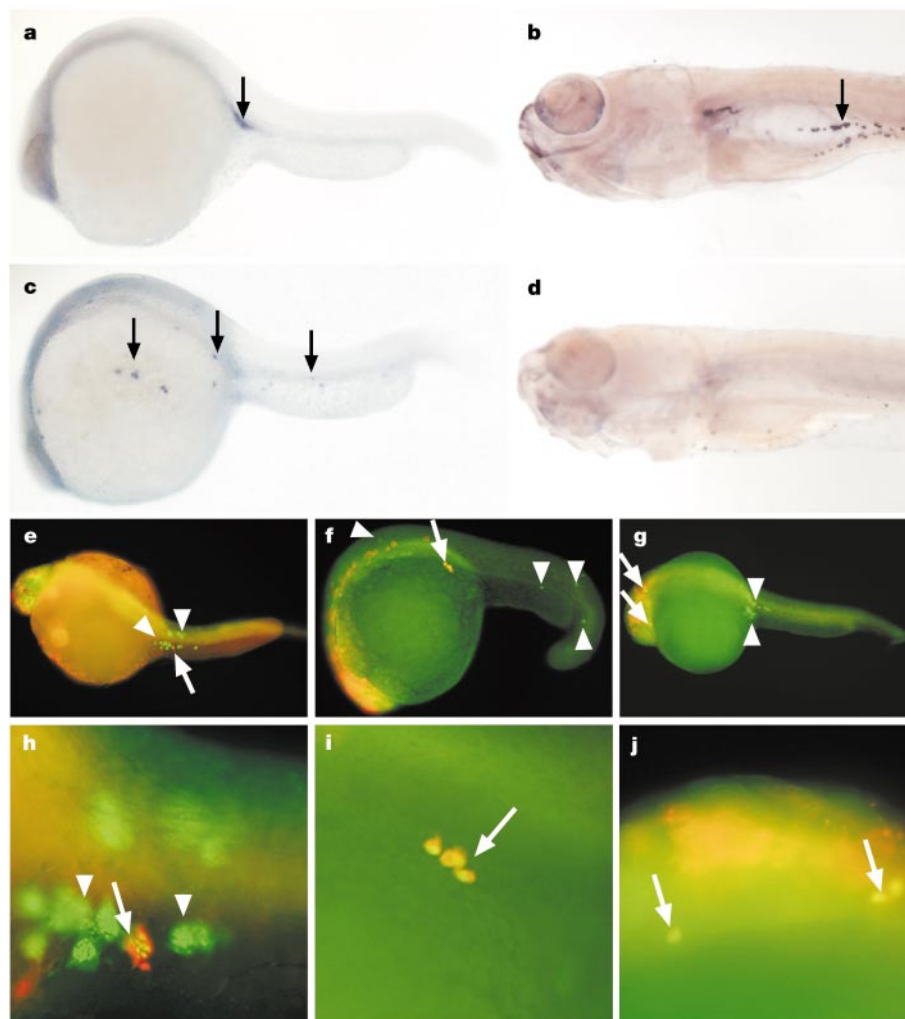


Figure 1 *odysseus* affects PGC positioning and is required within the germ line. **a–d**, Whole-mount immunostaining with an antibody against Vasa⁶ labels PGCs (arrows) in wild-type (**a**, **b**) and *odysseus* mutant embryos (**c**, **d**). At 30 h after fertilization, PGCs cluster at the anterior yolk extension in wild-type embryos (**a**), whereas in *odysseus* mutants (**c**), they are dispersed throughout the body. At 9 days after fertilization, PGCs have coalesced with the gonadal tissue in wild type (**b**), whereas in *odysseus* mutants no Vasa-positive PGCs are detectable (**d**). **e–j**, Transplantation of wild-type and *odysseus* mutant PGCs into wild-type and *odysseus* mutant embryos. One-cell-stage donor embryos were injected with biotin-dextran, and at the

8,000-cell stage approximately 100 donor cells were transplanted into recipient embryos of an equivalent stage. PGCs were identified by Vasa antibody staining (green) and donor-derived cells by anti-biotin antibody labelling (red). Transplantation of wild-type PGCs into wild-type embryos (**e**, magnified view in **h**) does not affect their ability to reach the anterior yolk extension. *odysseus* mutant PGCs transplanted into wild-type embryos (**g**, magnified view in **j**) still go astray, whereas wild-type PGCs transplanted into *odysseus* mutant embryos (**f**, magnified view in **i**) arrive at their target. Arrows indicate donor-derived PGCs; arrowheads indicate recipient-derived PGCs.

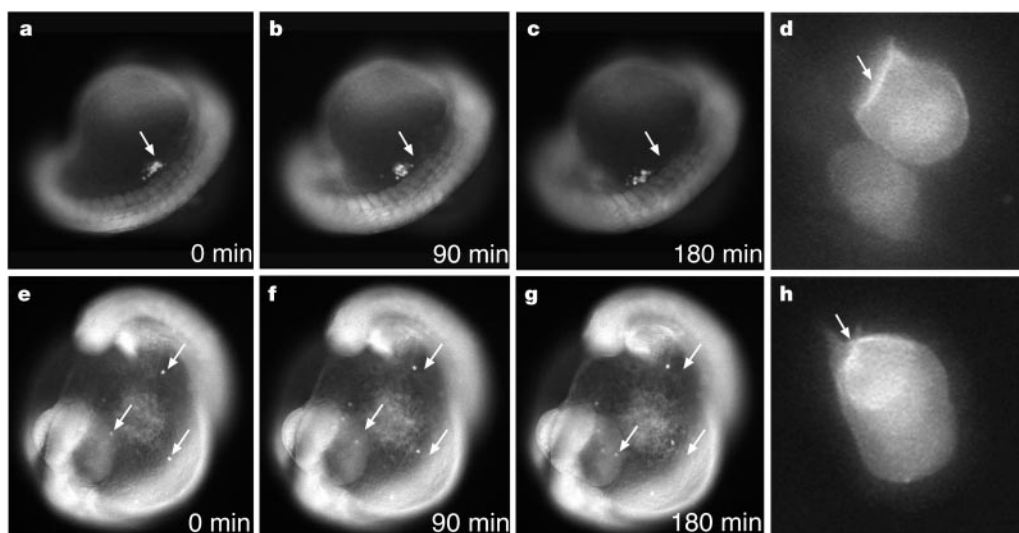


Figure 2 PGC migration in wild-type and *ody* mutant embryos. Expression of GFP⁹ (**a–c**, **e–g**) or GFP fused to a PH domain^{12,13} (**d**, **h**) was targeted to PGCs by fusing the coding sequences to the zebrafish *vasa* 3' untranslated region (UTR)¹⁶. Capturing consecutive images of individual embryos shows that PGCs (arrows indicate initial PGC position)

cluster and migrate in a posterior direction in wild-type embryos (**a–c**, anterior is to the top right). In *ody* mutant embryos (**e–g**, anterior is to the bottom left) PGCs (arrows) do not cluster, and migrate individually to ectopic sites. Germ cells recruit PH–GFP to locally restricted sites at the membrane (arrows, **d**, **h**) both in wild-type and *ody* mutants.

caudally to the anterior yolk extension, the future site of the gonads (Fig. 1a, b). During migration, PGCs are repelled from entering the trunk by an unknown signal emanating from the pronephros, but they are attracted by the anterior yolk extension^{8,9}. Among 1,358 genomes screened, we identified one strictly zygotic mutation, *odysseus* (*ody*), that resulted in incorrect positioning of PGCs without affecting general embryo morphology. Instead of being localized to the gonadal anlage, PGCs in *ody* mutants fail to cluster and are found randomly dispersed throughout the body at 30 h after fertilization (Fig. 1c).

To determine whether *ody* acts within the PGC or in the somatic tissue we generated genetic mosaics. On transplantation into *ody* mutant embryos, wild-type PGCs migrate to the correct position

(*n* = 9, Fig. 1f, i). However, mutant PGCs transplanted into wild-type embryos do not reach their target (*n* = 3, Fig. 1g, j). This indicates that the *ody* gene product is required within the PGCs for proper migration, in contrast to previously identified factors controlling PGC migration that act exclusively in somatic tissues^{2–4}.

The random positioning of PGCs in *ody* mutants could simply be the result of generally immotile cells being unable to leave their randomized sites of origin. Alternatively, *ody* might specifically affect the response of PGCs to chemotactic signals. *In vivo* imaging of embryos using a PGC-specific green fluorescent protein (GFP) marker⁹ revealed that mutant PGCs (Fig. 2e–g; see also Supplementary movie 2) display a similar degree of overall motility to wild-type cells (Fig. 2a–c; see also Supplementary movie 1). However,

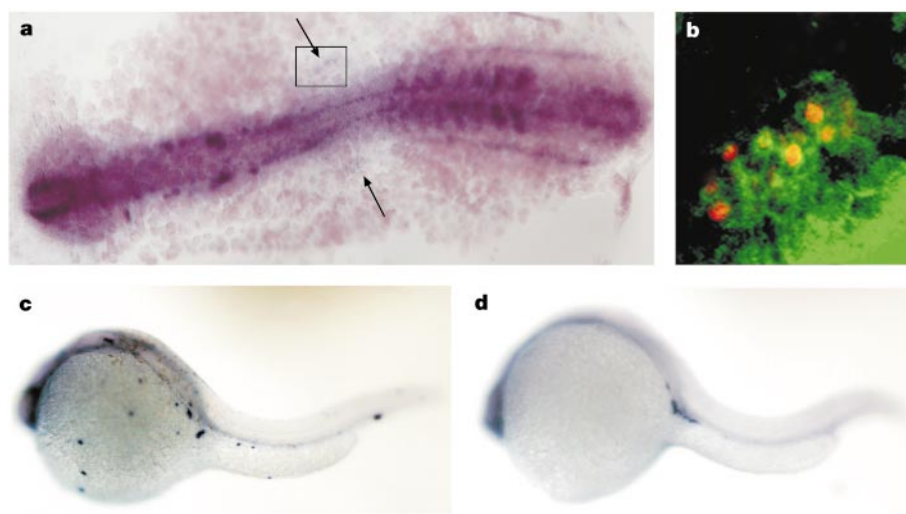


Figure 3 The G-protein-coupled chemokine receptor Cxcr4b is required for PGC guidance in zebrafish. **a**, **b**, *In situ* hybridization reveals that at the seven-somite stage Cxcr4b is, among other tissues¹⁵, expressed in two lateral clusters at the level of the second somite (arrows in **a**; the boxed area in **a** is magnified in **b**). Co-staining for Cxcr4b messenger RNA (green) and Vasa protein (red) shows that these two lateral clusters

correspond to PGCs (**b**). **c**, Embryos injected at the one-cell stage with Cxcr4b antisense morpholinos (0.2 mM, 5'-AGTGTGCTCAAAAAGGCGCAATAAG-3') display scattered PGCs (revealed by anti-Vasa staining) at 30 h after fertilization and thus recapitulate the *ody* phenotype. **d**, Injection of mismatch Cxcr4b morpholinos (0.2 mM, 5'-AGAGTCCTGAAAAAGGCGGAAAAAG-3') has no effect on PGC migration.

ody mutant PGCs fail to respond to both attractive and repulsive regions, and individual cells show faulty migration to various locations in the embryo, indicative of defective chemotaxis.

In many migratory cells, chemoattractants are sensed by G-protein-coupled receptors that signal through phosphatidylinositol-3-OH kinase (PI(3)K) to recruit pleckstrin homology (PH) domain-containing proteins to the leading edge (reviewed in refs 10, 11). Therefore, we investigated the subcellular localization of a PH-GFP fusion protein^{12,13} in PGCs *in vivo*. In wild-type PGCs, membrane recruitment of PH-GFP is locally restricted to the site of lamellipodium protrusion and remains relatively stably positioned over a longer period of time (Fig. 2d; see also Supplementary movie 3). Similarly, *ody* mutant PGCs recruit PH-GFP locally to the leading edge, but the site of recruitment changes position frequently (Fig. 2h; see also Supplementary movie 4). This indicates that *ody* mutant PGCs possess a functional machinery for establishing localized PI(3)K signalling and execution of the downstream migratory programme; however, they fail to maintain stably localized PI(3)K activation and thus lose directionality. This suggests that *ody* mutant PGCs do not perceive guidance cues. In mouse and zebrafish, PGCs migrate in close association¹ and *ody* seems to be required for this clustering in zebrafish. Although clustering might be a secondary effect of individual PGCs migrating to a common target, it is difficult to see how this might promote the intimate cell-cell contact observed between individual wild-type PGCs.

In principle, the *ody* mutation might affect PGC migration only indirectly by interfering with the normal PGC developmental programme. We find, however, that in 46% ($n = 78$) of the homozygous *ody* mutant embryos the gonads are partially populated with one to four PGCs and that a similar fraction of homozygous adult fish are fertile. Thus, *ody* mutant PGCs are, apart from the migration defect, normal with respect to their developmental programme and function.

We mapped *ody* (2,000 meioses) to a genomic interval of less than 200 kilobases (kb), 50.9 cM from the top of linkage group 9. Synteny and genome sequence analysis yielded two candidate genes, *arhgef4* and *cxc4b*. Sequencing demonstrated that the G-protein-coupled receptor *cxc4b* bears a nonsense mutation (Lys 239 to stop), deleting the carboxy terminal part of the third intracellular loop and the last two transmembrane domains. As these domains have been shown to be required generally for G-protein-coupled receptor signalling¹⁴, this mutation probably results in a complete loss of Cxcr4b protein function. *cxc4b* is expressed in various tissues¹⁵ and *in situ* hybridization shows that it is also expressed in migrating PGCs (Fig. 3a, b). Targeting the expression of wild-type *cxc4b* to PGCs¹⁶ fully rescues the *ody* mutant and injection of *cxc4b* antisense morpholino oligonucleotides reproduces the *ody* mutant phenotype (Fig. 3c, d), showing that *cxc4b* is the gene compromised in *ody* mutants. *cxc4b* belongs to a subfamily of G-protein-coupled chemokine receptors, members of which have, among other processes, been implicated in leukocyte migration¹⁷. Human CXCR4, known as fusin, forms, together with CD4, a receptor complex for HIV-1 (ref. 18). Mutant mice devoid of *Cxcr4* function are defective in B-cell lymphopoiesis, neuronal migration and vascularization^{19,20}. The comparatively subtle phenotype of *ody/cxc4b* mutants may be due to the presence of two Cxcr4 receptors with potentially split function in zebrafish, such that disruption of *cxc4b* reveals only a subset of the defects seen in mouse. This is supported by their exclusive expression patterns¹⁵. Alternatively, there may be variation between zebrafish and mouse in the roles of the Cxcr4 receptors.

As cell culture experiments and loss of function analysis in mouse have implicated the chemokine SDF-1 as the ligand for Cxcr4 (refs 21–23), we analysed a zebrafish homologue of SDF-1. SDF-1 is expressed in a dynamic pattern that prefigures the route of PGC migration (Fig. 4a–d). Reducing SDF-1 activity by antisense morpholinos results in incorrect positioning of PGCs as in *ody*

mutant embryos (28 out of 85) (Fig. 4e, g). Moreover, ectopic expression of SDF-1 in wild-type embryos recruited between 10% and 30% of the endogenous PGCs to positions near the SDF-1-expressing cells to which they would never normally migrate (15 out of 23) (Fig. 4f). This response is receptor-dependent as SDF-1-overexpressing cells failed to attract *ody* mutant PGCs to ectopic positions (0 out of 15) (Fig. 4h).

From these results we conclude that the attractant SDF-1 and its receptor *ody/Cxcr4b* comprise a system for guiding vertebrate PGCs throughout their entire migration. Furthermore, this finding reveals a role for G-protein-coupled receptors in embryonic development—other roles include the mediation of attractive chemotaxis of *Dictyostelium* amoebae, leukocytes^{10,11} and, more recently, neurons²⁴. Notably, PGCs with inactive SDF-1 or *cxc4b* also ignore repulsive signals. Although this might simply be due to the fact that the early faulty migration of PGCs means that they never encounter such signals, an alternative explanation would be that SDF-1 and *cxc4b* are also mediating repulsion, either by a repellent directly

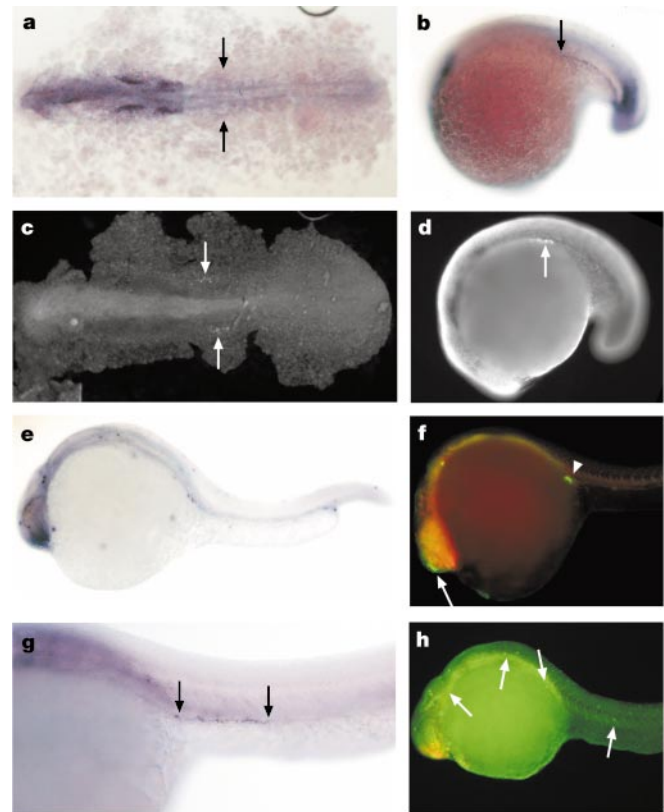


Figure 4 SDF-1 is a chemoattractant for PGCs. Zebrafish SDF-1 RNA is expressed in a dynamic pattern that demarcates the prospective PGC migration route (a–d, g). **a, c**, In a dorsal view, *in situ* hybridization with a SDF-1 probe (**a**) and immunostaining against Vasa protein (**c**) show that SDF-1 is expressed proximal to the two PGC clusters (arrows) at the level of the second somite in a seven-somite-stage embryo. **b, d**, In a 17-somite-stage embryo, the anterior boundary of the SDF-1 expression domain (**b**, arrow) reaches to the level of the eighth somite where the PGC cluster resides (**d**, arrow). **g**, At 36 h after fertilization the SDF-1 expression domain is restricted to the final target of PGCs (arrows). **e**, Embryos injected with SDF-1 antisense morpholinos (2 mM, 5'-CGCTACTACTTTGCTATCCATGCCA-3') display scattered PGCs at 30 h after fertilization, a phenotype indistinguishable from the *ody* mutant (compare with Fig. 1c). SDF-1 mismatch morpholinos (2 mM, 5'-CGGTATTACTTTCTATCCTTGGA-3') do not affect PGC migration (data not shown). **f**, Transplantation of blastula cells from embryos injected with *in vitro*-transcribed SDF-1 mRNA and biotin-dextran into wild-type embryos attracts PGCs (green) towards SDF-1-expressing cells (red). **h**, Transplanting similarly treated cells into *ody* mutant embryos has no effect on PGC distribution. Arrows indicate incorrectly positioned PGCs; arrowheads indicate correctly positioned PGCs (**f, h**).

interfering with the system or by attractant-activated processes being a prerequisite for responding to repulsion. Given the length and complexity of the migration route in zebrafish, it seems unlikely that PGC guidance relies solely on the SDF-1/Cxcr4b pair. When considering the multiple sites of SDF-1 expression, it is likely that either post-transcriptional modifications of SDF-1 or co-expression of repellents may assist PGC navigation. In *Drosophila*, two enzymes involved in lipid metabolism, Wunen and Columbus^{2–4}, affect PGC migration presumably by controlling production of unknown repulsive or attractive signals. It will be interesting to see how such signals might tie in with the guidance system identified here and whether a widely conserved system exists for guiding PGCs to their gonadal targets. □

Received 8 November; accepted 2 December 2002; doi:10.1038/nature01338.
Published online 18 December 2002.

- Wylie, C. Germ cells. *Cell* **96**, 165–174 (1999).
- Starz-Gaiano, M., Cho, N. K., Forbes, A. & Lehmann, R. Spatially restricted activity of a *Drosophila* lipid phosphatase guides migrating germ cells. *Development* **128**, 983–991 (2001).
- Van Doren, M., Broihier, H. T., Moore, L. A. & Lehmann, R. HMG-CoA reductase guides migrating primordial germ cells. *Nature* **396**, 466–469 (1998).
- Zhang, N., Zhang, J., Purcell, K. J., Cheng, Y. & Howard, K. The *Drosophila* protein Wunen repels migrating germ cells. *Nature* **385**, 64–67 (1997).
- Knauf, H., Pelegri, F., Bohnmann, K., Schwarz, H. & Nusslein-Volhard, C. Zebrafish vasa RNA but not its protein is a component of the germ plasm and segregates asymmetrically before germline specification. *J. Cell Biol.* **149**, 875–888 (2000).
- Yoon, C., Kawakami, K. & Hopkins, N. Zebrafish vasa homologue RNA is localized to the cleavage planes of 2- and 4-cell-stage embryos and is expressed in the primordial germ cells. *Development* **124**, 3157–3165 (1997).
- Olsen, L. C., Aasland, R. & Fjose, A. A vasa-like gene in zebrafish identifies putative primordial germ cells. *Mech. Dev.* **66**, 95–105 (1997).
- Weidinger, G. et al. Regulation of zebrafish primordial germ cell migration by attraction towards an intermediate target. *Development* **129**, 25–36 (2002).
- Knauf, H., Steinbeisser, H., Schwarz, H. & Nusslein-Volhard, C. An evolutionary conserved region in the vasa 3' UTR targets RNA translation to the germ cells in the zebrafish. *Curr. Biol.* **12**, 454–466 (2002).
- Chung, C. Y., Potikyan, G. & Firtel, R. A. Control of cell polarity and chemotaxis by Akt/PKB and PI3 kinase through the regulation of PAKα. *Mol. Cell* **7**, 937–947 (2001).
- Iijima, M., Huang, Y. E. & Devreotes, P. Temporal and spatial regulation of chemotaxis. *Dev. Cell* **3**, 469–478 (2002).
- Meili, R. et al. Chemoattractant-mediated transient activation and membrane localization of Akt/PKB is required for efficient chemotaxis to cAMP in *Dictyostelium*. *EMBO J.* **18**, 2092–2105 (1999).
- Parent, C. A., Blacklock, B. J., Froehlich, W. M., Murphy, D. B. & Devreotes, P. N. G protein signaling events are activated at the leading edge of chemotactic cells. *Cell* **95**, 81–91 (1998).
- Wess, J. G-protein-coupled receptors: molecular mechanisms involved in receptor activation and selectivity of G-protein recognition. *FASEB J.* **11**, 346–354 (1997).
- Chong, S. W., Emelyanov, A., Gong, Z. & Korzh, V. Expression pattern of two zebrafish genes, *cxcr4a* and *cxcr4b*. *Mech. Dev.* **109**, 347–354 (2001).
- Wolke, U., Weidinger, G., Koprinner, M. & Raz, E. Multiple levels of posttranscriptional control lead to germ line-specific gene expression in the zebrafish. *Curr. Biol.* **12**, 289–294 (2002).
- Horuk, R. Chemokine receptors. *Cytokine Growth Factor Rev.* **12**, 313–335 (2001).
- Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872–877 (1996).
- Zou, Y. R., Kottmann, A. H., Kuroda, M., Taniuchi, I. & Littman, D. R. Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development. *Nature* **393**, 595–599 (1998).
- Tachibana, K. et al. The chemokine receptor CXCR4 is essential for vascularization of the gastrointestinal tract. *Nature* **393**, 591–594 (1998).
- Bleul, C. C. et al. The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature* **382**, 829–833 (1996).
- Oberlin, E. et al. The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nature* **382**, 833–835 (1996).
- Nagasawa, T. et al. Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature* **382**, 635–638 (1996).
- Zhu, Y. et al. Role of the chemokine SDF-1 as the meningeal attractant for embryonic cerebellar neurons. *Nature Neurosci.* **5**, 719–720 (2002).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank F. Schnorrer for discussions and critical comments on the manuscript. The PH-GFP fusion protein was a gift from M. Landree and P. Devreotes. H.K. was in part supported by a Boehringer Ingelheim predoctoral fellowship. R.G. was supported by the German Human Genome Project (DHGP). This work was supported by the Max-Planck Gesellschaft. Publication costs were in part defrayed by the Fonds der Chemischen Industrie.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.K.
(e-mail: knaut@saturn.med.nyu.edu).

The Tübingen 2000 Screen Consortium

E. Busch-Nentwich*, R. Dahm*, H.-G. Frohnhofer*, H. Geiger*, D. Gilmour*, S. Holley*, J. Hooge*, D. Jülich*, H. Knaut*, F. Maderspacher*, H.-M. Maischein*, C. Neumann*, T. Nicolson*, C. Nüsslein-Volhard*, H. Roehl*, U. Schönberger*, C. Seiler*, C. Söllner*, M. Sonawane*, F. van Bebber*, A. Wehner*, C. Weiler*, P. Erkert*, H. Habeck*, U. Hagner*, C. E. Hennen Kapst*, A. Kirchner*, T. Koblizek*, U. Langheinrich*, C. Loeschke*, C. Metzger*, R. Nordin*, J. Odenthal*, M. Pezzuti*, K. Schlombst*, J. deSantana-Stamm*, T. Trowet*, G. Vacun*, B. Walderich*, A. Walkert*, C. Weiler†

*Max Planck Institut für Entwicklungsbiologie, Abteilung III/Genetik, †Artemis Pharmaceuticals GmbH, an Exelixis company, Spemannstrasse 35, 72076 Tübingen, Germany

SAP is required for generating long-term humoral immunity

Shane Crotty*, Ellen N. Kersh*, Jennifer Cannons†, Pamela L. Schwartzberg† & Rafi Ahmed*

*Emory Vaccine Center and Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta, Georgia 30322, USA

†National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

Long-lived plasma cells and memory B cells are the primary cellular components of long-term humoral immunity and as such are vitally important for the protection afforded by most vaccines. The SAP gene has been identified as the genetic locus responsible for X-linked lymphoproliferative disease, a fatal immunodeficiency^{1–4}. Mutations in SAP have also been identified in some cases of severe common variable immunodeficiency disease^{5,6}. The underlying cellular basis of this genetic disorder remains unclear. We have used a SAP knockout mouse model system to explore the role of SAP in immune responses. Here we report that mice lacking expression of SAP generate strong acute IgG antibody responses after viral infection, but show a near complete absence of virus-specific long-lived plasma cells and memory B cells, despite the presence of virus-specific memory CD4⁺ T cells. Adoptive transfer experiments show that SAP-deficient B cells are normal and the defect is in CD4⁺ T cells. Thus, SAP has a crucial role in CD4⁺ T-cell function: it is essential for late B-cell help and the development of long-term humoral immunity but is not required for early B-cell help and class switching.

The B-cell immune response defects in SAP[−] mice were examined using lymphocytic choriomeningitis virus (LCMV). Adult mice infected with LCMV (Armstrong strain) make vigorous primary immune responses, clear the infection within a week and maintain strong virus-specific T-cell and B-cell memory for life^{7–10}. To characterize the kinetics of the anti-LCMV B-cell response, wild-type and SAP[−] mice were infected with LCMV, and numbers of virus-specific antibody-secreting cells (ASCs) were measured in the spleen. SAP[−] mice made a substantial LCMV-specific ASC response that peaked at almost wild-type amounts on day 8 after infection (Fig. 1a). Antibody-secreting cells were also enumerated by cytometry, and at day 8 after infection a large population of immunoglobulin-γ (IgG)⁺CD138⁺ plasma cells was seen in both wild-type and SAP[−] mice (Fig. 1c). In wild-type mice, the number of splenic LCMV-specific ASCs dropped after viral clearance but then stabilized at about 10% of the peak amount by day 15 after infection (Fig. 1a). This concentration of about 10,000 LCMV-specific ASCs per spleen remains as a stable population for life (ref. 9 and Fig. 1a). In contrast, SAP[−] mice showed a marked loss of LCMV-specific ASCs after viral clearance, which dropped to nearly undetectable amounts by day 15 (≤300 per spleen).

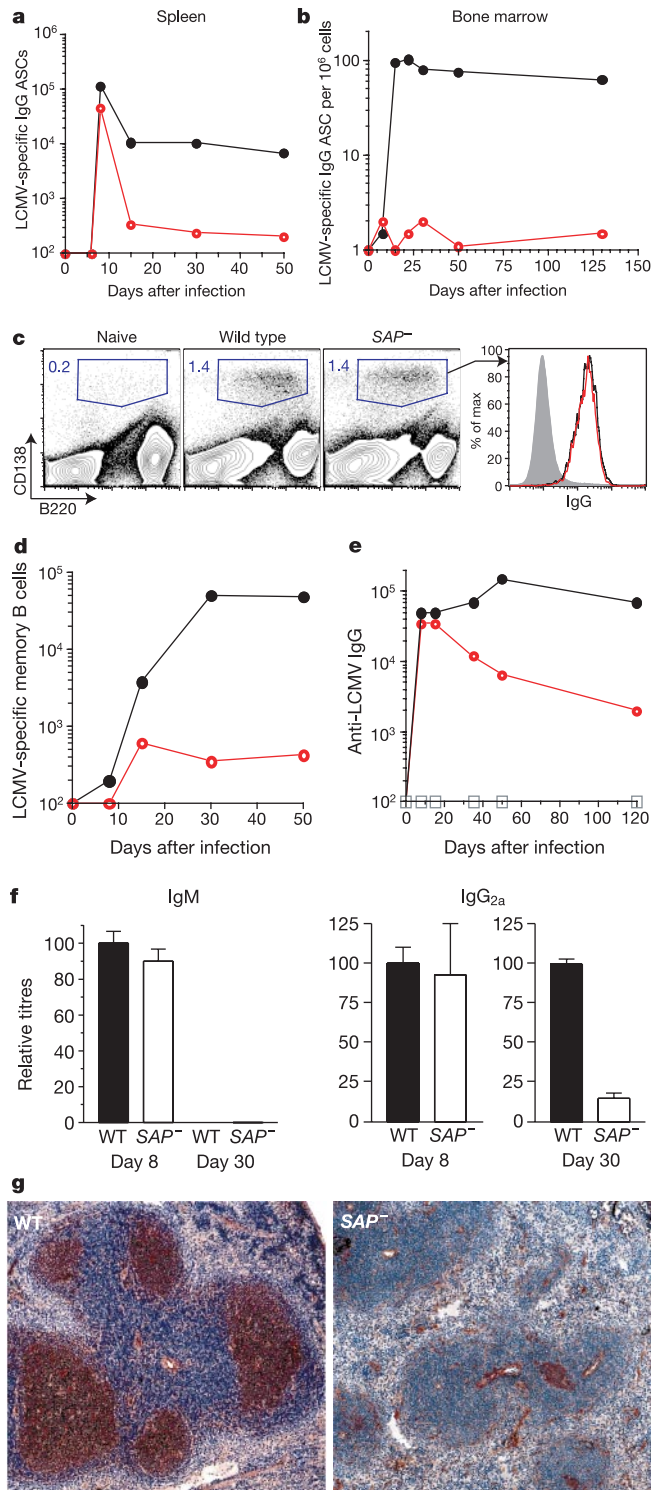


Figure 1 Defective long-term humoral immune responses in $SAP^{-/-}$ mice. **a**, LCMV-specific IgG antibody-secreting cells (ASCs) in the spleen were tracked in wild-type (black) and $SAP^{-/-}$ (red) mice infected with LCMV. Responses in each group generally varied less than twofold. **b**, Long-lived plasma cell responses were measured in bone marrow. **c**, Plasma cells in spleen were also measured at day 8 by cytometry (CD138⁺B220⁺; left) and were more than 98% IgG⁺ in both wild-type and $SAP^{-/-}$ mice (right). Wild-type, black; $SAP^{-/-}$, red; negative control, grey. **d**, LCMV-specific memory B cells in the spleen. Most data points represent the average of 3–6 mice. **e**, Kinetics of the anti-LCMV IgG antibody response in sera of wild-type and $SAP^{-/-}$ mice measured by ELISA. Naive mouse serum is shown in grey (squares). **f**, Isotype profiles of the anti-LCMV IgM and IgG_{2a} antibody responses. Wild type, black; $SAP^{-/-}$, white. **g**, Germinal centres at day 16. PNA⁺ germinal centre B cells are brown, haematoxylin counterstain is blue.

Long-lived plasma cells are located primarily in the bone marrow and are the dominant source of virus-specific antibody production in immune animals^{9,11–13}. The population of LCMV-specific ASCs in the bone marrow increased in wild-type mice until day 30 and thereafter was stably maintained (Fig. 1b). LCMV-immune $SAP^{-/-}$ mice had a severe deficit of virus-specific plasma cells in the bone marrow, containing 50-fold fewer cells than wild-type mice (Fig. 1b).

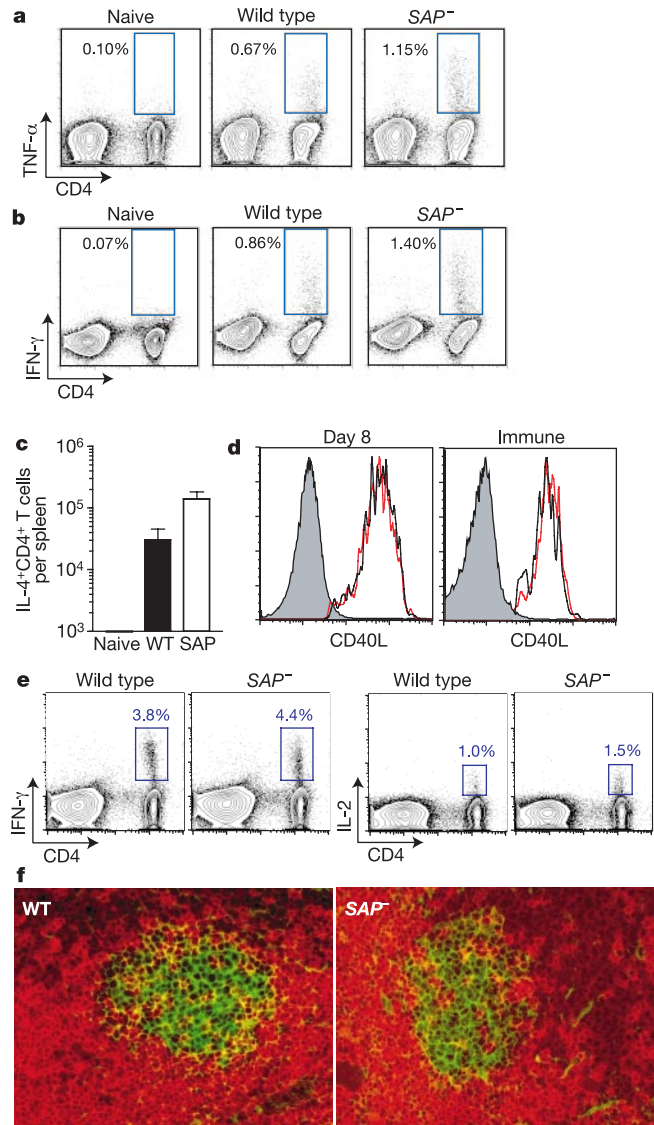


Figure 2 CD4⁺ T-cell functional analysis. Day 8 after infection, splenocytes were stimulated for 5 h with LCMV gp61–80 MHC class II peptide in the presence of brefeldin A and then analysed by intracellular cytokine staining. **a**, TNF- α ⁺CD4⁺. **b**, INF- γ ⁺CD4⁺. **c**, LCMV-specific IL-4⁺ CD4⁺ T-cell numbers were also increased in $SAP^{-/-}$ mice at day 8, as measured by ELISPOT. **d**, CD40L expression was normal in LCMV-specific CD4⁺ T cells after stimulation with gp61–80 peptide and standard intracellular staining for CD40L. Day 8 after infection and immune animals (day 90) are shown. Plots are gated on INF- γ ⁺CD4⁺ cells. Wild type, black; $SAP^{-/-}$, red; unstimulated, naive CD4⁺ T cells, grey. **e**, Numbers of LCMV-specific CD4⁺ T cells were normal in $SAP^{-/-}$ knockout mice at day 30 after infection, as measured by intracellular cytokine staining (INF- γ , top; IL-2, bottom) of CD4⁺ T cells specific for the LCMV gp61–80 peptide. The percentages shown are the number of gated cytokine-positive CD4⁺ T cells out of the total number of CD4 cells. **f**, CD4⁺ T-cell localization in germinal centres of WT and $SAP^{-/-}$ mice at day 16 after infection. Anti-CD4 staining in red, PNA⁺ germinal centre staining in green.

Although the acute ASC response was close to normal in $SAP^{-/-}$ mice, the production of long-lived plasma cells was clearly defective. To examine whether this humoral memory compartment defect in $SAP^{-/-}$ mice was limited to the production of long-lived plasma cells, we quantified the number of LCMV-specific memory B cells in $SAP^{-/-}$ mice. Wild-type mice began to produce LCMV-specific memory B cells by day 8 after infection, and a stable population of 50,000 LCMV-specific memory B cells per spleen was generated by day 30 (Fig. 1d). In contrast, $SAP^{-/-}$ mice had no detectable memory B cells at day 8, and at day 30 there were only about 300 LCMV-specific memory B cells per spleen (Fig. 1d). That population did not increase in size at later time points (Fig. 1d), leaving $SAP^{-/-}$ mice with a marked 100-fold reduction in virus-specific memory B cells. This phenotype was not due to a generalized loss of memory lymphocytes, because normal or increased numbers of memory $CD8^{+}$ T cells^{14,15} and $CD4^{+}$ T cells (see Fig. 2) were observed in $SAP^{-/-}$ mice.

The kinetics of the antiviral IgG antibody response was examined in $SAP^{-/-}$ mice to determine the relationship between the serum antibody concentrations and ASCs. High titres of antibodies against LCMV were maintained in wild-type mice, whereas titres of IgG antibody peaked in $SAP^{-/-}$ mice at day 15 and thereafter declined (Fig. 1e). Thus, serum IgG antibody titres reflected the plasma cell kinetics.

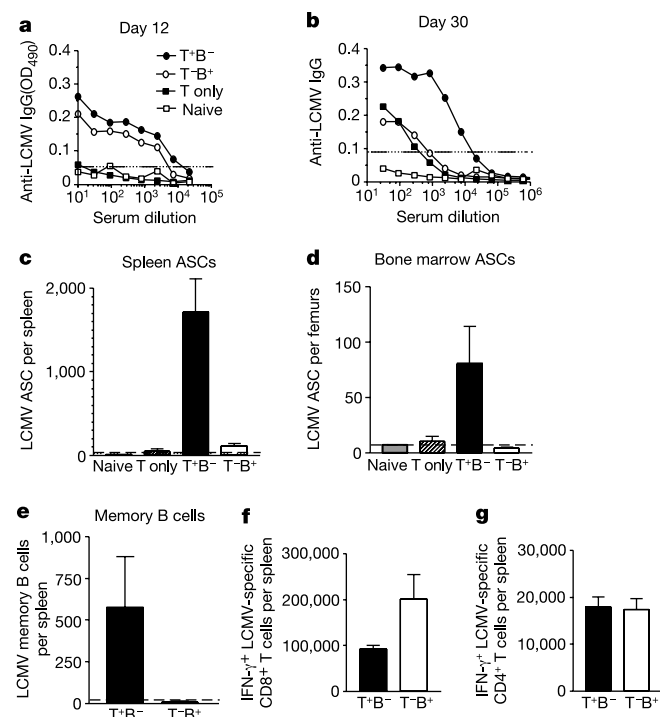


Figure 3 SAP expression is not required in B cells for long-term humoral memory. A set of adoptive transfers involved transferring wild-type T lymphocytes alone (T only), wild-type T lymphocytes mixed with $SAP^{-/-}$ B lymphocytes (T⁺B⁻); or $SAP^{-/-}$ T lymphocytes and wild-type B lymphocytes (T⁺B⁺). A total of 50×10^6 total cells was transferred into naive irradiated (600 rad) wild-type recipients, which were then infected with LCMV. **a**, Acute IgG anti-LCMV antibody responses were made in both the T⁺B⁻ and T⁺B⁺ transfers, as measured by ELISA at day 12 after infection. **b–g**, Immune mice were then killed at day 32 after infection and serum antibody titres (**b**), virus-specific ASCs in spleen (**c**), virus-specific long-lived plasma cells in bone marrow (**d**), virus-specific memory B cells (**e**), virus-specific memory CD8⁺ T cells (np396-epitope-specific; **f**), and virus-specific memory CD4⁺ T cells (gp61-80-epitope-specific; **g**) were measured. A marked loss of memory B cell and long-lived plasma cell responses was observed in mice receiving $SAP^{-/-}$ T cells. Broken line indicates limit of detection.

In CD40L-deficient mice and humans, a severe defect in the generation of antigen-specific CD4⁺ T-helper (T_H) cells leaves B cells unable to switch isotypes. This results in hyper-IgM syndrome^{16,17}. To examine whether SAP deficiency affects isotype switching, isotypes were examined during both the acute and immune phases after infection. Anti-LCMV IgM titres in $SAP^{-/-}$ mice were normal at day 6, and IgG concentrations were substantial (70% of wild type) at day 8 (Fig. 1e, f). No anti-LCMV IgM was detectable at day 30 in wild-type or $SAP^{-/-}$ mice (Fig. 1f). This was consistent with ASC ELISPOT results (data not shown). The predominant IgG subclass in a normal anti-LCMV response is IgG_{2a}, with minor amounts of IgG_{2b}, IgG₃, and IgG₁. No significant difference in the amount of IgG_{2a} between wild-type and $SAP^{-/-}$ mice was observed at day 8, although substantially less IgG_{2a} was detected at day 30 in $SAP^{-/-}$ mice, as expected (Fig. 1f). In contrast, IgG₁ antibodies to LCMV were not detectable in $SAP^{-/-}$ mice at any time point (data not shown).

The presence of a strong, acute B-cell response and the lack of long-lived plasma cells and memory B cells implied a defect in germinal centre formation or maintenance. We therefore examined germinal centres by histological assessment. A severe reduction in the number and size of germinal centres was obvious in the $SAP^{-/-}$ mice at all time points after infection (day 16, Fig. 1g; days 8, 22, 30 and 36, data not shown). Germinal centres peaked at about day 16 after infection in wild-type mice, at which point there were 2,000 germinal centre B cells per mm². In $SAP^{-/-}$ mice, this population was reduced roughly 10-fold. The basic anatomic organization of splenic CD4⁺ T cells and B220⁺ B cells was generally comparable in wild-type and $SAP^{-/-}$ mice (data not shown).

Are the long-lived plasma cell and memory B cell defects in $SAP^{-/-}$ mice caused by defective B cells or T cells? SAP is an Src homology domain 2 (SH2) domain protein that is expressed in most T and natural killer (NK) cells⁴ and may be expressed in certain subsets of

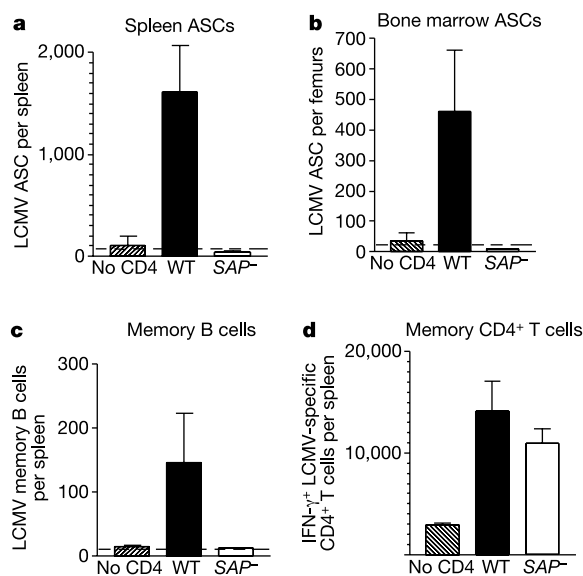


Figure 4 SAP expression is required in CD4⁺ T_H cells to generate long-term humoral immunity. A set of adoptive transfers involved transferring wild-type CD4⁺ T cells (WT), $SAP^{-/-}$ CD4⁺ T cells ($SAP^{-/-}$) or no CD4⁺ T cells (no CD4), along with wild-type B cells, CD8⁺ T cells and other splenocytes into irradiated (620 rad) recipient mice. A total of 50×10^6 cells was transferred per mouse, and the recipient mice were infected with LCMV. Mice were killed at day 32 after infection and virus-specific ASCs in spleen (**a**), virus-specific long-lived plasma cells in bone marrow (**b**), virus-specific memory B cells (**c**) and virus-specific memory CD4⁺ T cells (gp61-80-epitope-specific; **d**) were measured. Broken line indicates limit of detection.

B cells^{18,19}. SAP can bind to the cytoplasmic tails of several surface proteins of the SLAM family (SLAM, 2B4, CD84, Ly9), which are expressed in T, NK, dendritic and/or B cells^{2,4,18,20}. SLAM, in particular, is highly expressed in both T and B cells. We therefore explored the ability of $SAP^{-/-}$ B cells to proliferate and to differentiate in response to different stimuli. *In vitro* proliferation of $SAP^{-/-}$ B cells and differentiation into ASCs was equivalent to wild-type under all conditions tested (data not shown).

Genetic deficiencies of molecules such as CD40L, CD28 and B7.1 or B7.2 that disrupt B-cell help generally have significantly reduced, or absent, $CD4^{+}$ T_H -cell responses. In $SAP^{-/-}$ mice, by contrast, antigen-specific $CD8^{+}$ and $CD4^{+}$ T cells are generally functional and are present in normal to increased amounts^{14,15}. In fact, the overall magnitude of the LCMV-specific $CD4^{+}$ T-cell response in $SAP^{-/-}$ mice was usually higher than in wild-type mice (refs 14, 15, and Fig. 2). Frequencies of virus-specific tumour-necrosis factor- α (TNF- α^{+}), interleukin 2 (IL-2)⁺ and interferon- γ (IFN- γ)⁺ $CD4^{+}$ T cells were all normal or increased at day 8 (Fig. 2a, b, and data not shown) and were generally normal, if not higher than in wild type, in immune $SAP^{-/-}$ mice (Fig. 2e). Unexpectedly, given the apparent T_H1 predisposition of $SAP^{-/-}$ $CD4^{+}$ T cells^{14,15}, frequencies of IL-4⁺ $CD4^{+}$ T cells were also increased in $SAP^{-/-}$ mice at day 8 and were normal in $SAP^{-/-}$ immune mice (Fig. 2c and data not shown). CD40L expression, which is crucial for B-cell help^{17,21}, also seemed to be normal in antigen-specific $SAP^{-/-}$ $CD4^{+}$ T cells at day 8 after infection and in immune animals (Fig. 2d). In addition, we observed a strong antiviral acute B-cell response with a dominant IgG component, indicating that $CD4^{+}$ T-cell help was functional.

Migration of $CD4^{+}$ T cells is important for providing appropriate B-cell help^{22–26}. Although the germinal centres themselves were infrequent and irregular in shape in $SAP^{-/-}$ mice, the frequency of splenic CXCR5⁺ $CD4^{+}$ T cells was normal in $SAP^{-/-}$ mice at day 8 after infection (data not shown), and the frequency of $CD4^{+}$ T cells in and around the few germinal centres observed in $SAP^{-/-}$ mice was comparable to that in wild-type mice (Fig. 2f).

Because it was not apparent which cell type was responsible for the severe long-term humoral immunity defects observed in $SAP^{-/-}$ mice, we carried out a series of adoptive transfer experiments. B and T lymphocytes from wild-type and $SAP^{-/-}$ mice were mixed and adoptively transferred into naive irradiated wild-type mice. Wild-type B cells and $SAP^{-/-}$ T cells were transferred into one set of mice ($T^{-}B^{+}$), and $SAP^{-/-}$ B cells and wild-type T cells were transferred into a separate set of mice ($T^{+}B^{-}$). The mice were infected with

LCMV and their immune responses were measured. Irradiated mice that received no adoptive transfer and were then infected with LCMV made no LCMV-specific T-cell or B-cell responses (data not shown). To control specifically for the endogenous B-cell response in the irradiated recipients, a set of naive mice received adoptive transfers of wild-type T cells alone and were then infected with LCMV. Those mice made negligible B-cell responses (Fig. 3a–d).

Serum antibody concentrations in mice receiving the adoptive transfers were measured at day 12 after infection to confirm that both experimental groups of mice ($T^{-}B^{+}$ and $T^{+}B^{-}$) made acute anti-LCMV IgG responses, whereas control mice did not (Fig. 3a). Immune mice were then tested for the presence of serum antibodies against LCMV, as well as LCMV-specific ASCs in spleen, long-lived plasma cells in bone marrow and memory B cells. Mice receiving $SAP^{-/-}$ B cells made and maintained strong LCMV-specific humoral immune responses (Fig. 3b–e). However, mice receiving $SAP^{-/-}$ T cells and wild-type B cells had marked defects in long-term humoral immunity, with barely detectable LCMV-specific long-lived plasma cell or memory B-cell responses (Fig. 3b–e). LCMV-specific $CD4^{+}$ and $CD8^{+}$ T-cell responses were present in both groups of mice, as expected (Fig. 3f, g).

To test whether SAP expression is required in $CD4^{+}$ T cells for generating long-term humoral immunity, we carried out a second set of adoptive transfer experiments. These mice all received wild-type APCs, $CD8^{+}$ T cells and B cells, and in addition received wild-type $CD4^{+}$ T cells, $SAP^{-/-}$ $CD4^{+}$ T cells or no $CD4^{+}$ T cells. Mice were infected with LCMV, and serum antibody titres were measured at day 15 to confirm that mice receiving either wild-type $CD4^{+}$ T cells or $SAP^{-/-}$ $CD4^{+}$ T cells both made acute anti-LCMV IgG responses (data not shown). Immune mice were then tested for the presence of LCMV-specific ASCs in spleen, long-lived plasma cells in bone marrow and memory B cells (Fig. 4a–c). Mice receiving wild-type $CD4^{+}$ T cells made and maintained strong LCMV-specific humoral immune responses (Fig. 4a–c). However, mice receiving $SAP^{-/-}$ $CD4^{+}$ T cells were severely defective in generating long-term humoral immunity, with no detectable LCMV-specific long-lived plasma cells or memory B cells (Fig. 4a–c). Notably, the magnitude of the LCMV-specific $CD4^{+}$ T-cell response was comparable in both the wild-type and $SAP^{-/-}$ transfers (Fig. 4d).

SAP expression is not required in cell types other than $CD4^{+}$ T cells for generating long-term humoral immunity, because adoptive transfer of wild-type $CD4^{+}$ T cells in conjunction with $SAP^{-/-}$ non- $CD4^{+}$ splenocytes rescued the humoral immunity phenotype to that of wild type (Fig. 5a). In addition, we determined that the $SAP^{-/-}$ $CD4^{+}$ T_H -cell loss of function is a recessive phenotype, not a dominant or gain-of-function phenotype, because adoptive transfer of an equal mix of wild-type and $SAP^{-/-}$ $CD4^{+}$ T cells resulted in a wild-type humoral immune response phenotype (Fig. 5b).

Expression of SAP in $CD4^{+}$ T cells is therefore essential for generating long-lived plasma cells and memory B cells. Additional in-depth studies will be required to identify the important effector, developmental or migratory functions affected by SAP signalling in $CD4^{+}$ T cells. A genetic defect resulting in $CD4^{+}$ T cells competent for early B-cell help but not competent for late B-cell help is unusual. People with SAP mutations frequently present with low IgG antibody titres, but many different immunological abnormalities are seen, which are often complicated by the presence of chronic infections^{2,27,28}. Many deaths from X-linked lymphoproliferative (XLP) disease are associated with uncontrolled Epstein–Barr virus (EBV) infections. Our data suggest that an inability to sustain anti-EBV humoral immune responses may contribute to this phenotype. Therapy with intravenous immunoglobulin can be effective in some cases of XLP, indicating the possible value of a sustained humoral response against EBV. In addition, many individuals with XLP are plagued by severe chronic or recurring bacterial infections. Again, this may be due to an inability to develop effective long-term humoral immunity.

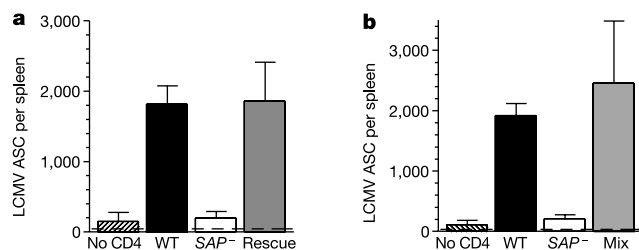


Figure 5 SAP expression requirements for generating long-term humoral immunity. **a**, Adoptive transfer of wild-type $CD4^{+}$ T cells along with $SAP^{-/-}$ non- $CD4^{+}$ splenocytes (B cells, $CD8^{+}$ T cells and other $CD4^{-}$ splenocytes) into irradiated (620 rad) recipient mice ('Rescue') was carried out to determine whether SAP expression is required in non- $CD4^{+}$ types of cell. Adoptive transfers of wild-type $CD4^{+}$ T cells (WT), $SAP^{-/-}$ $CD4^{+}$ T cells ($SAP^{-/-}$) or no $CD4^{+}$ T cells (no CD4), along with wild-type B cells, $CD8^{+}$ T cells, and other splenocytes were done as controls. **b**, Adoptive transfer of wild-type and $SAP^{-/-}$ $CD4^{+}$ T cells together ('Mix') into irradiated mice receiving wild-type non- $CD4^{+}$ splenocytes was done to determine whether $SAP^{-/-}$ T cells are dominant or recessive for the block in long-term humoral immunity.

Although an encompassing description of XLP is made difficult by the heterogeneity of the human phenotypes, many SAP-deficient people have low concentrations of IgG and frequently make marginal IgG antibody responses to human vaccines^{6,27,28}. Notably, of the subset of individuals with common variable immunodeficiency identified to have SAP mutations, most show serious immunoglobulin deficiencies^{5,6}. We have shown here that SAP has a central role in the development of immunological memory. SAP expression in CD4⁺ T cells is required to drive the production of memory B cells and long-lived plasma cells; but, unexpectedly, SAP expression is not required for initial B-cell help or isotype switching. The important role of SAP in humoral immune responses suggests that manipulation of SAP-mediated signalling pathways may have therapeutic benefits. □

Methods

Mice and virus infections

Wild-type 129 Sv/Ev/Tac mice were purchased from Taconic and were used as control mice in all experiments excluding the adoptive transfers. C57BL/6J (B6) mice were purchased from the Jackson Laboratory and were used in the adoptive transfer experiments. SAP^{−/−} mice were generated as described¹⁴ and maintained on a 129 Sv/Ev/Tac background (in all experiments excluding the adoptive transfers) or backcrossed for six generations to B6 (for adoptive transfer experiments). We carried out genotyping as described¹⁴. Adult male mice were used in all experiments. The Armstrong CA 1371 strain of LCMV was used, and virus stocks were grown and quantified as described²⁹. All mouse infections were done by intraperitoneal injection with 2×10^5 plaque-forming units of LCMV Armstrong.

Cell preparation and flow cytometry

Single-cell suspensions of spleen and lymph node were prepared by standard gentle mechanical disruption. We isolated femoral bone marrow as described⁹.

Surface staining for flow cytometry or histology used monoclonal antibodies to CD4, CD8, CD3, B220, CD44, CD138, CXCR5 and IgD (BD Pharmingen); polyclonal goat antibodies against mouse IgG₁ (Caltag Laboratories) and fluorescein isothiocyanate (FITC)-labelled peanut agglutinin (PNA) (Vector Laboratories). For histology, biotinylated primary antibodies were recognized with streptavidin-conjugated Alexa-568 (Molecular Probes). Intracellular cytokine and CD40L staining was done by standard techniques after a 5-h stimulation⁷, using antibodies against IFN- γ , TNF- α , IL-2 and CD40L from BD Pharmingen. Flow cytometry samples were acquired on a FACSCalibur instrument (Beckton Dickinson) and analysed using FlowJo software (TreeStar).

Plasma cell ELISPOT and T-cell ELISPOT

LCMV-specific plasma cells were quantified by a modification of a described ELISPOT method^{9,11}. Lysate from LCMV-infected Vero cells was used as antigen. We used biotinylated goat antibodies against mouse IgG₁, IgM, IgG_{2a}, IgG_{2b} or IgG₃ (Caltag Laboratories) as detecting antibodies. Biotinylated antibodies against mouse IgG₁ was obtained from BD Pharmingen. Horseradish peroxidase (HRP)-conjugated Avidin D was from Vector Laboratories. We carried out IL-4 ELISPOT assay for gp61-antigen-specific CD4⁺ T cells as described⁸.

Quantification of memory B cells

Virus-specific memory B cells were measured by a modification of a described limiting-dilution method³⁰. We cultured splenocytes for 6 d in a flat-bottomed 96-well dish in the presence of 1×10^6 irradiated (1,200 rad) feeder splenocytes, 0.4 μ g of R595 lipopolysaccharide (Alexis Biochemicals) and 20 μ l of T-cell- and B-cell-stimulating concanavalin A supernatant (IGEN) in R-10 medium (RPMI-1640, 100 IU ml^{−1} penicillin, 100 IU ml^{−1} streptomycin and 10% fetal calf serum), supplemented with freshly added 50 μ M β -mercaptoethanol, in a total volume of 200 μ l. Three-fold dilutions of splenocytes were tested in replicates of 12 wells each. After 6 d of polyclonal activation, cells were washed and transferred to LCMV-antigen-coated 96-well Multiscreen-HA filter plates (Millipore) and ASC ELISPOTs were done as described above. We calculated the frequency of LCMV-specific memory B cells by standard limiting-dilution single-hit kinetics.

Enzyme-linked immunosorbent assay

LCMV-specific serum antibody titres were determined by a solid-phase direct ELISA as described²⁹, using HRP-conjugated goat antibodies against mouse Ig of the isotype or IgG subclass of interest (Caltag Laboratories) except IgG₁, for which a BD Pharmingen biotinylated antibody was used. We defined naive mouse serum titres as zero.

Lymphocyte purification

The adoptive transfers shown in Fig. 3 used splenocytes and lymph-node cells purified on the basis of B220 expression. We carried out two rounds of positive selection using MACS beads conjugated with antibodies against B220 (Miltenyi). After positive column selection, B-cell purity was 96–99%, as determined by flow cytometry surface phenotyping, with less than 0.5% contaminating CD4⁺ T cells. T cells were collected as the

flow through from the first B220-positive selection column. T-cell preparations contained roughly 50% CD4⁺ T cells, 30% CD8⁺ cells and 0.5–1% B cells. SAP^{−/−} lymphocytes were obtained from SAP^{−/−} B6 male mice, and wild-type lymphocytes were obtained from SAP^{+/+} B6 littermates. Mixed populations were reconstituted as 50% B cells and 50% non-B cells.

The adoptive transfers shown in Figs 4 and 5 used splenocytes and lymph-node cells purified on the basis of CD4 expression. We carried out one round of positive selection using MACS beads conjugated with antibodies against CD4 (Miltenyi). After positive column selection, CD4⁺ T-cell purity was 90–95%, as determined by flow cytometry. Non-CD4 cell preparations were collected as the flow through of the column and contained roughly 70% B cells, 15–25% CD8⁺ T cells and 7–10% other cell types, including less than 0.3% CD4⁺ T cells. SAP^{−/−} lymphocytes were obtained from SAP^{−/−} B6 male mice, and wild-type lymphocytes were obtained from B6 males. Mixed populations were reconstituted as 20–25% CD4⁺ T cells and 75–80% non-CD4 cells. In the 'Mix' experiment, 12.5×10^6 wild-type CD4⁺ T cells, 12.5×10^6 SAP^{−/−} CD4⁺ T cells and 37.5×10^6 wild-type non-CD4⁺ splenocytes were transferred to provide a full amount of both CD4⁺ T-cell groups.

Irradiation and adoptive transfer

Naive B6 mice received 600–620 rad total body irradiation to deplete the lymphocyte populations. Mice were reconstituted 12–18 h later with 50×10^6 lymphocytes of interest by intravenous injection. LCMV injections were done immediately after cell transfers.

Received 22 July; accepted 14 November 2002; doi:10.1038/nature01318.

- Coffey, A. J. *et al.* Host response to EBV infection in X-linked lymphoproliferative disease results from mutations in an SH2-domain encoding gene. *Nature Genet.* **20**, 129–135 (1998).
- Morra, M. *et al.* X-linked lymphoproliferative disease: a progressive immunodeficiency. *Annu. Rev. Immunol.* **19**, 657–682 (2001).
- Nichols, K. E. *et al.* Inactivating mutations in an SH2 domain-encoding gene in X-linked lymphoproliferative syndrome. *Proc. Natl Acad. Sci. USA* **95**, 13765–13767 (1998).
- Sayos, J. *et al.* The X-linked lymphoproliferative-disease gene product SAP regulates signals induced through the co-receptor SLAM. *Nature* **395**, 462–469 (1998).
- Morra, M. *et al.* Alterations of the X-linked lymphoproliferative disease gene SH2D1A in common variable immunodeficiency syndrome. *Blood* **98**, 1321–1325 (2001).
- Nistala, K. *et al.* X-linked lymphoproliferative disease: three atypical cases. *Clin. Exp. Immunol.* **126**, 126–130 (2001).
- Murali-Krishna, K. *et al.* Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. *Immunity* **8**, 177–187 (1998).
- Whitmire, J. K., Asano, M. S., Murali-Krishna, K., Suresh, M. & Ahmed, R. Long-term CD4 Th1 and Th2 memory following acute lymphocytic choriomeningitis virus infection. *J. Virol.* **72**, 8281–8288 (1998).
- Slifka, M. K., Antia, R., Whitmire, J. K. & Ahmed, R. Humoral immunity due to long-lived plasma cells. *Immunity* **8**, 363–372 (1998).
- Homann, D., Teyton, L. & Oldstone, M. B. Differential regulation of antiviral T-cell immunity results in stable CD8⁺ but declining CD4⁺ T-cell memory. *Nature Med.* **7**, 913–919 (2001).
- Slifka, M. K., Matloubian, M. & Ahmed, R. Bone marrow is a major site of long-term antibody production after acute viral infection. *J. Virol.* **69**, 1895–1902 (1995).
- Manz, R. A., Thiel, A. & Radbruch, A. Lifetime of plasma cells in the bone marrow. *Nature* **388**, 133–134 (1997).
- O'Connor, B. P., Cascalho, M. & Noelle, R. J. Short-lived and long-lived bone marrow plasma cells are derived from a novel precursor population. *J. Exp. Med.* **195**, 737–745 (2002).
- Czar, M. J. *et al.* Altered lymphocyte responses and cytokine production in mice deficient in the X-linked lymphoproliferative disease gene SH2D1A/DSHP/SAP. *Proc. Natl Acad. Sci. USA* **98**, 7449–7454 (2001).
- Wu, C. *et al.* SAP controls T cell responses to virus and terminal differentiation of TH2 cells. *Nature Immunol.* **2**, 410–414 (2001).
- Durandy, A. & Honjo, T. Human genetic defects in class-switch recombination (hyper-IgM syndromes). *Curr. Opin. Immunol.* **13**, 543–548 (2001).
- Banchereau, J. *et al.* The CD40 antigen and its ligand. *Annu. Rev. Immunol.* **12**, 881–922 (1994).
- Shlapatska, L. M. *et al.* CD150 association with either the SH2-containing inositol phosphatase or the SH2-containing protein tyrosine phosphatase is regulated by the adaptor protein SH2D1A. *J. Immunol.* **166**, 5480–5487 (2001).
- Tangye, S. G., van de Weert, B. C. M., Avery, D. T. & Hodgkin, P. D. CD84 is up-regulated on a major population of human memory B cells and recruits the SH2 domain containing proteins SAP and EAT-2. *Eur. J. Immunol.* **32**, 1640–1649 (2002).
- Latour, S. *et al.* Regulation of SLAM-mediated signal transduction by SAP, the X-linked lymphoproliferative gene product. *Nature Immunol.* **2**, 681–690 (2001).
- van Essen, D., Kikutani, H. & Gray, D. CD40 ligand-transduced co-stimulation of T cells in the development of helper function. *Nature* **378**, 620–623 (1995).
- Forster, R. *et al.* CCR7 coordinates the primary immune response by establishing functional microenvironments in secondary lymphoid organs. *Cell* **99**, 23–33 (1999).
- Ansel, K. M., McHeyzer-Williams, L. J., Ngo, V. N., McHeyzer-Williams, M. G. & Cyster, J. G. *In vivo*-activated CD4 T cells upregulate CXC chemokine receptor 5 and reprogram their response to lymphoid chemokines. *J. Exp. Med.* **190**, 1123–1134 (1999).
- Campbell, D. J., Kim, C. H. & Butcher, E. C. Separable effector T cell populations specialized for B cell help or tissue inflammation. *Nature Immunol.* **2**, 876–881 (2001).
- Kim, C. H. *et al.* Subspecialization of CXCR5⁺ T cells: B helper activity is focused in a germinal center-localized subset of CXCR5⁺ T cells. *J. Exp. Med.* **193**, 1373–1381 (2001).
- MacLennan, I. C. Germinal centers. *Annu. Rev. Immunol.* **12**, 117–139 (1994).
- Grierson, H. L., Skare, J., Hawk, J., Pauza, M. & Purtilo, D. T. Immunoglobulin class and subclass deficiencies prior to Epstein-Barr virus infection in males with X-linked lymphoproliferative disease. *Am. J. Med. Genet.* **40**, 294–297 (1991).
- Sumegi, J. *et al.* Correlation of mutations of the SH2D1A gene and Epstein-Barr virus infection with

clinical phenotype and outcome in X-linked lymphoproliferative disease. *Blood* **96**, 3118–3125 (2000).

29. Ahmed, R., Salmi, A., Butler, L. D., Chiller, J. M. & Oldstone, M. B. Selection of genetic variants of lymphocytic choriomeningitis virus in spleens of persistently infected mice. Role in suppression of cytotoxic T lymphocyte response and viral persistence. *J. Exp. Med.* **160**, 521–540 (1984).
30. Slifka, M. K. & Ahmed, R. Limiting dilution analysis of virus-specific memory B cells by an ELISPOT assay. *J. Immunol. Methods* **199**, 37–46 (1996).

Acknowledgements We thank L. Mijares for technical assistance, and J. Gibson Lanier and D. Barber for expertise. This work was supported by intramural and extramural NIH funding. S.C. is a Cancer Research Institute postdoctoral fellow.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.A. (e-mail: ra@microbio.emory.edu).

Responses of ferns to red light are mediated by an unconventional photoreceptor

Hiroko Kawai*, Takeshi Kanegae*, Steen Christensen*†, Tomohiro Kiyosue†‡, Yoshikatsu Sato*‡, Takato Imaizumi*‡, Akeo Kadota* & Masamitsu Wada*†

* Department of Biological Sciences, Graduate School of Science, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan

† Division of Biological Regulation and Photobiology, National Institute of Basic Biology, Okazaki, Aichi 444-8585, Japan

Efficient photosynthesis is essential for plant survival. To optimize photosynthesis, plants have developed several photoresponses. Stems bend towards a light source (phototropism), chloroplasts move to a place of appropriate light intensity (chloroplast photorelocation) and stomata open to absorb carbon dioxide. These responses are mediated by the blue-light receptors phototropin 1 (phot1) and phototropin 2 (phot2) in *Arabidopsis* (refs 1–5). In some ferns, phototropism and chloroplast photorelocation are controlled by red light as well as blue light⁶. However, until now, the photoreceptor mediating these red-light responses has not been identified. The fern *Adiantum capillus-veneris* has an unconventional photoreceptor, phytochrome 3 (phy3), which is a chimera of the red/far-red light receptor phytochrome and phototropin⁷. We identify here a function of phy3 for red-light-induced phototropism and for red-light-induced chloroplast photorelocation, by using mutational analysis and complementation. Because phy3 greatly enhances the sensitivity to white light in orienting leaves and chloroplasts, and *PHY3* homologues exist among various fern species, this chimaeric photoreceptor may have had a central role in the divergence and proliferation of fern species under low-light canopy conditions.

We screened 17 strains of *rap* (red-light aphototropic) mutants from the fern *Adiantum capillus-veneris*, including five previously reported strains⁸, and found all to be deficient both in phototropism and chloroplast relocation under red light in protonemal cells, whereas the same responses were normal under blue light. These two red light effects in wild-type protonemata can be reversed by far-red light, implicating a phytochrome as the responsible photo-

receptor^{9,10}. Because *Adiantum* leaves (sporophytic) have been reported to show both phototropic bending¹¹ and chloroplast relocation¹² in response to red light, we tested whether the deficiency in protonemal cells (gametophytic) also exists for the sporophyte in a *rap2* mutant. We found that the *rap2* mutant is also deficient for both of these physiological responses in leaves (Fig. 1). Thus the same component is involved in these red-light responses

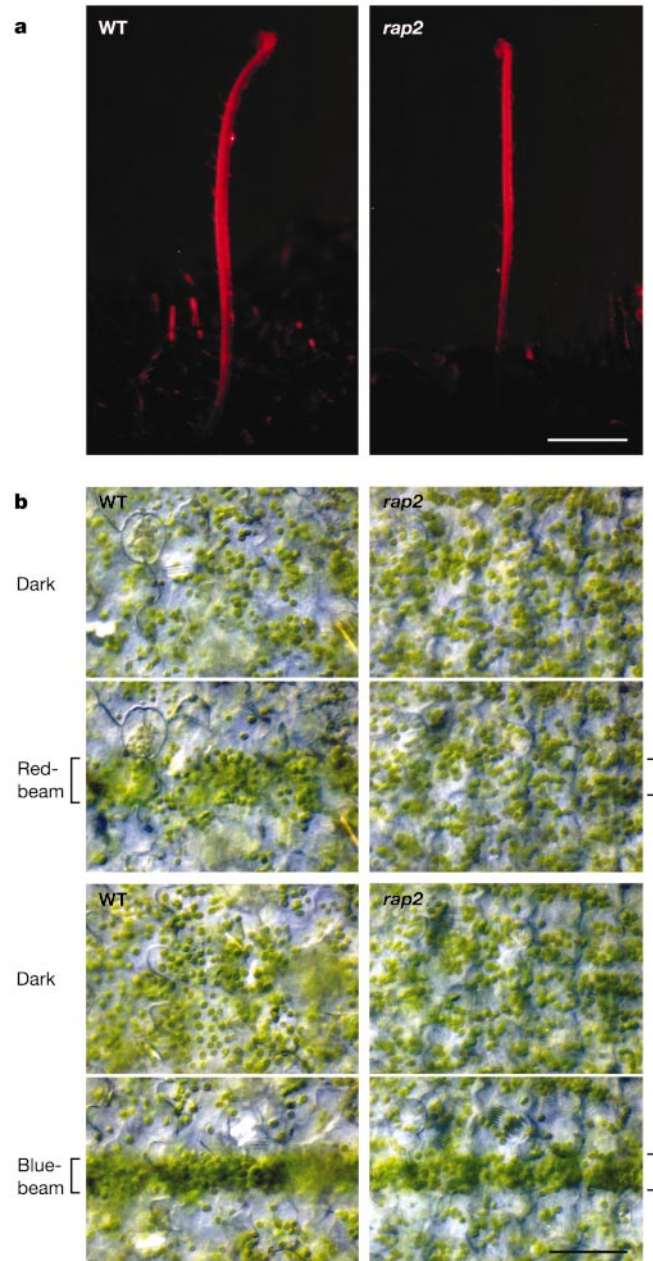


Figure 1 *rap2* mutants lack red-light-induced phototropism and chloroplast relocation movement in the sporophyte generation. **a**, Red-light-induced phototropic response in young etiolated leaves of wild-type (WT) and *rap2* sporophytes. The leaves were unilaterally irradiated for 4.5 h with red light ($0.22 \mu\text{mol m}^{-2} \text{s}^{-1}$) from the right. The scale bar represents 5 mm. **b**, Chloroplast relocation movement in leaflet cells of wild-type (WT) and *rap2* sporophytes. After taking a photograph of chloroplast distribution in darkness (upper images), the leaflets were further incubated for 3 h in the dark, and then partly irradiated for 3 h with bands of red ($0.55 \mu\text{mol m}^{-2} \text{s}^{-1}$) or blue ($0.38 \mu\text{mol m}^{-2} \text{s}^{-1}$) light from a 20- μm width microbeam, as indicated by the brackets on both sides of the panels, before taking a second photograph. The scale bar represents 50 μm .

‡ Present addresses: Division of Speciation Mechanisms 2, National Institute of Basic Biology, Okazaki, Aichi 444-8585, Japan (Y.S.); Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA (T.I.); Gene Research Center, Kagawa University, Miki-cho Kita-gun, Kagawa 761-0795, Japan (T.K.).

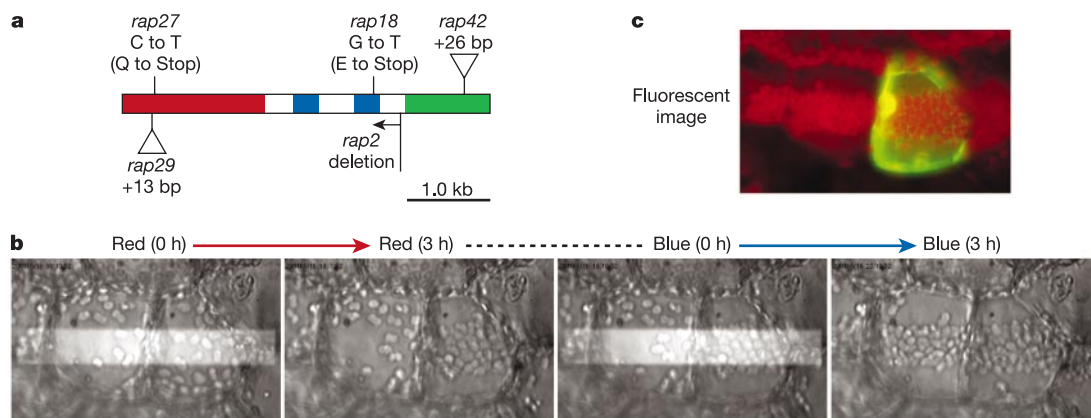


Figure 2 Functional correlation of phy3 with the *rap* phenotype. **a**, Positions of *PHY3* mutations in *rap* mutants (for details see text). Red, blue and green bars represent phytochrome chromophore-binding, LOV1 and LOV2, and serine/threonine protein kinase domains respectively. **b**, Serial images of chloroplast relocation movements in *rap2* mutant cells into which the *PHY3* gene was introduced (right cell) or was not (left cell). A band across the cells, indicated by the lighter area, was irradiated continuously with a

20- μm width microbeam of red light ($0.55 \mu\text{mol m}^{-2} \text{s}^{-1}$). After 3 h, the microbeam was switched to blue light ($0.38 \mu\text{mol m}^{-2} \text{s}^{-1}$) and irradiated for an additional 3 h. **c**, The green fluorescence image in the right cell (taken immediately after the **b** series) indicates GFP expression as a marker for gene transfer. The red fluorescence arises from chlorophyll.

both in gametophytes and sporophytes. A possible candidate for the mutated gene is that encoding the red-light photoreceptor involved in these responses.

In *Adiantum*, three genes encoding phototropin-related proteins have been identified. Phot1 (ref. 13) and phot2 (T. Kagawa, personal communication) are conventional phototropin proteins, containing two LOV (light, oxygen or voltage) domains and a serine/threonine kinase domain. However, phy3 is a chimaeric protein with the chromophore-binding domain of phytochrome at its amino-terminal end and a full-length phototropin at its carboxy-terminal end⁷. Phy3 binds phycocyanobilin (which can substitute for the native phytochrome chromophore, phytochromobilin) and the bound form is red/far-red photoreversible, demonstrating that phy3 acts as a phytochrome⁷. The LOV domains from phy3 were found to bind flavin mononucleotide (FMN), as is the case for higher-plant phototropins^{14,15}, suggesting that phy3 may absorb blue light as well as red/far-red light. Considering that the red/far-red-reversible responses in the *rap* mutants are defective, whereas the blue-light responses are normal, and considering that two phototropins known to mediate phototropism and chloroplast movements in *Arabidopsis*⁴ are present, it appeared possible that deleterious mutations at the *PHY3* locus may have caused the loss in red-light sensitivity noted above. Hence, we investigated the *PHY3* genomic sequences of five ethyl methanesulfonate (EMS)-generated *rap* mutants, and found that all five mutants showed defects in the *PHY3* gene (Fig. 2a). The mutations are as follows: deletion of the entire region upstream from base 3319 in *rap2*; nonsense mutations at positions 3052 and 349 in *rap18* and *rap27*, respectively; and duplication of base pairs 305–317 and 4075–4100, both of which cause frame-shift mutations, in *rap29* and *rap42*, respectively. These results strongly suggest that the mutations in *PHY3* cause the mutant phenotype, and that phy3 is the photoreceptor for both phototropism and chloroplast relocation in response to red light.

To test the function of phy3, we did a complementation test with the *rap2* mutant. The wild-type *PHY3* gene and a green fluorescent protein (GFP) marker gene were co-transformed into *rap2* prothallial cells by particle bombardment for transient expression. In a GFP-expressing *rap2* cell (the right cell in Fig. 2b, c), also considered to be expressing *PHY3*, chloroplasts accumulated in the region irradiated with a red-light microbeam whereas they failed to gather into the irradiated region in an untransformed cell (the left cell in Fig. 2b, c) (see Supplementary Information Movie 1). By contrast,

chloroplast accumulation was induced only by blue light and not by red light in cells transformed with GFP only (see Supplementary Information Movie 2). The complementation of red-light-induced chloroplast movement in *rap2* by the *PHY3* gene clearly demonstrates that the photoreceptor phy3 mediates this response. Because all of the *rap* mutants lack both red-light-mediated phototropism and chloroplast movement, it is highly likely that phy3 is also the photoreceptor for red-light-induced phototropism. A direct test of this hypothesis must await stable transformation in *Adiantum*.

Wild-type *Adiantum* has a functional *PHY3* gene in addition to *PHOT1* and *PHOT2* genes. Hence, it can use both red and blue light to mediate phototropism and chloroplast relocation. If the effects of these three photoreceptors are additive, they then provide an indirect test of a role for phy3 in the sporophyte generation. One would expect plants with all three photoreceptors to have a lower threshold for phototropism in response to unilateral white light than plants lacking one of the three, and to be generally less sensitive. We compared sensitivities of the phototropic responses of young leaves of *rap2* sporophytes with the sensitivities of young leaves of wild-type plants under low white-light conditions (7.8×10^{-4} to $4.0 \times 10^{-3} \mu\text{mol m}^{-2} \text{s}^{-1}$) (Fig. 3). The *rap2* mutant showed severely impaired photosensitivity in low light in contrast to the wild type. Similar experiments were done with *rap18* and *rap29* with similar results (data not shown). The loss of red-light sensitivity in phy3-deficient mutants (Fig. 1) supports the contention that the photosensory function of phy3 substantially enhances the phototropic response in *Adiantum* to white light.

Previous results with *Adiantum* gametophytes showed that chloroplast photorelocation did not occur in either sub-threshold red light or sub-threshold blue light. However, if sub-threshold red and blue lights are given sequentially in either order chloroplast relocation can be induced¹⁶, suggesting that signals from both red and blue light receptors work additively. Under canopy shade, although this light environment is depleted of red with respect to far-red, the red/far-red ratios are still sufficient to maintain effective steady-state levels of the physiologically active form of phytochrome (Pfr)¹⁷. Although the light-induced signalling mechanism of phy3 has yet to be solved, the capacity to couple the photosensitivity of phytochrome to phototropin kinase activity¹⁴ would allow both red and blue light to function in phototropism and chloroplast movement, a distinct advantage under low-light conditions.

A phy3-like chimaeric photoreceptor is not found in *Arabidopsis* nor has it been reported in any plant other than *Adiantum*. Given

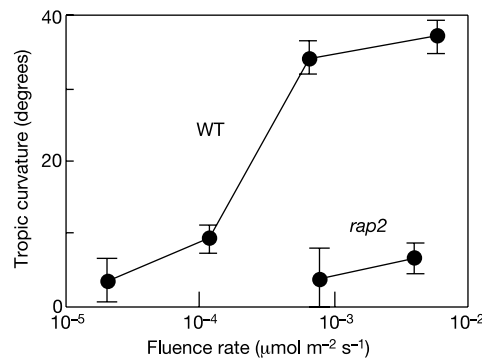


Figure 3 The phototropic response of *rap2* leaves is less sensitive to dim white light than that of wild-type (WT) leaves. Young etiolated leaves of *rap2* and wild-type plants were irradiated continuously from one side with white light of various fluence rates.

Curvatures developing during 4.5 h of irradiation are shown. Each point represents the mean $\pm$ the standard error with sample sizes between 7 and 47 individuals.

that many ferns such as *Dryopteris filix-mas*¹⁸, *D. sparsa*¹⁹, *Onoclea sensibilis*²⁰, *Pteridium aquilinum*²¹ and *Matteuccia struthiopteris* (data not shown) show red-light-induced phototropism and chloroplast movement, *phy3* may well be found in other fern species apart from *Adiantum*. To address this possibility we looked for *PHY3* sequences in several fern species. *PHY3*-homologous sequences were found in *D. filix-mas* (accession number AB089917), *O. sensibilis* (AB089919) and *Hypolepis punctata* (AB089918), all species within the 'polypodiaceous' ferns²². These species are considered highly derived in the fern phylogenetic tree^{23,24}. However, *PHY3*-homologous sequences could not be detected in the more primitive ferns such as *Osmunda japonica* (Osmundaceae) and *Lygodium japonicum* (Schizaeaceae). Because red-light-mediated chloroplast movement and phototropism could not be observed in the more primitive ferns *O. japonica* and *L. japonicum* (data not shown), it is likely that they all lack an expressed *PHY3* gene.

On the basis of the fossil record, the Osmundaceae (the oldest extant leptosporangiate fern family) arose during the Permian, the Schizaeaceae during the Triassic, and the Cyatheaceae during the Early Jurassic periods. Other more advanced fern families including the 'polypodiaceous' ferns appeared thereafter from the Jurassic to the Cretaceous, and most of the extant families can be traced from between the Upper Cretaceous and the Tertiary^{22,25}. Although ecological circumstances during the Palaeozoic and Mesozoic are not well understood, fossil records show us the flora of these eras, which help us to infer global physiognomic changes. Since the Permian, gymnosperms accelerated expansion of their distribution, but it was probably between the Triassic and Jurassic that the inland forest, with a well-developed canopy, was established worldwide and when major forests dominated by gymnosperms such as conifers and ginkgos flourished^{25,26}. This development suggests the wide establishment of a new niche 'under canopy' by the Jurassic.

The advent of dense forest, and the subsequent weak-light environment found below its canopy, probably presented a suitable niche for low-light-adapted ferns. The timing of the explosive proliferation of 'polypodiaceous' fern species may be coincident with the acquisition of the red-light-induced responses mediated by *phy3*. We propose that the appearance of *phy3* was a critical step for the proliferation of ferns, and has had an important role in the competitive success of past and present fern species. □

Methods

Mutant selection and phenotypic analysis

Mutagenesis of wild-type spores of *Adiantum* and selection of *rap* mutants was done as reported previously⁸. Sporophytes cultivated in the greenhouse of Tokyo Metropolitan University were used for measurement of the phototropic response. All existing leaves were removed and the plants were cultivated in darkness for about one month until

etiolated leaves developed. The leaves were then exposed to unilateral red or white light for 4.5 h. Phototropic curvature was measured as reported previously¹¹. Red light was obtained by passing light from a fluorescent tube (FL20SD, Toshiba Lighting and Technology) through a sheet of red plastic (Shinkolite A, No. 102; Mitsubishi Rayon). White light was obtained from fluorescent tubes (FL10D, Toshiba Lighting and Technology). Neutral-density filter (Polycolor No. 98, Tokyo Butai Showmei) and filter paper were used to attenuate the fluence rate, measured by using a DataLogger (LI-1000, Li-Cor) with a quantum sensor (LI-190SA, Li-Cor) or calculated from distances between the leaf and light source.

To follow light-induced chloroplast movement in sporophytes, young leaflets ranging from 3 to 4 mm wide were evacuated in a syringe filled with de-ionized water and transferred into a cuvette composed of a glass slide and a round coverslip, the latter supported by a ring-shaped silicon-rubber spacer. The cavity was filled with de-ionized water containing 0.1% agar (BA-30, INA), and incubated in darkness for one day. The area of the leaf to be irradiated was photographed under an Axioplan microscope (Zeiss) with a digital colour camera (FD-420 M, Flovel) immediately after removing the leaf from darkness. The leaf was then placed on the stage of the microbeam irradiator and kept in darkness for an additional 3 h. Microbeam irradiation was done as described previously²⁷.

Sequence analysis

Genomic DNA from *rap* gametophytes was extracted with a DNeasy Plant Mini Kit (Qiagen). The *PHY3* coding regions were amplified by polymerase chain reaction (PCR) using Platinum Pfx DNA polymerase (Invitrogen) and five sets of *PHY3* gene-specific primers. The PCR products were purified with a QIAquick PCR Purification Kit (Qiagen) and directly sequenced as described previously⁷.

Genomic DNA from fern species other than *H. punctata* was isolated by the cetyltrimethylammonium bromide (CTAB) method²⁸. *PHY3*-like sequence fragments from these ferns were amplified by PCR with *EX Taq* polymerase (Takara Bio). Degenerate primers used in first-round PCR were designed around highly conserved regions both in the phytochrome chromophore-binding domain (5'-GTIATGGYITWYMRITTCAYGARGA-3') and the phototropin serine/threonine kinase domain (5'-TTYTCIGGYTTIARRTCICKRTA-3'). PCR products were purified as described above and an aliquot of each was used for nested PCR with 5'-ATGGAYYTIGTIAARTGYGAYGG-3' and 5'-ACIARRTGACISWNCCNGTRTC-3' primers. Fragments of desired length were recovered by agarose gel electrophoresis and cloned into the pGEM-T easy plasmid vector (Promega), from which at least four independent clones were fully sequenced.

Transient assay for mutant complementation

A cauliflower mosaic virus (CaMV) 35S promoter-driven *PHY3* gene (pBI221-*PHY3*) was generated by subcloning the coding region of the *PHY3* sequence into the *Xba*I-*Sac*I sites of pBI221 (Clontech). A modified *GFP* gene driven by the CaMV 35S promoter (*psmRS-GFP*) was used as a marker for gene transfer. The pBI221-*PHY3* and/or the *psmRS-GFP* constructs were introduced into *rap2* prothalli by particle bombardment. Prothalli were first cut into pieces with a Polytron (PT 1200, Kinematica) and the fragments grown under white light (20–30 μmol m⁻² s⁻¹) for 1–4 weeks. They were then subjected to osmotic pretreatment by placing them on media containing 0.25 M mannitol for 1 h before bombardment with a Biolistic Particle Delivery System-100/He (Bio-Rad). Delivery conditions were as follows: gold particle diameter was 1.6 μm; helium pressure was 1350 p.s.i.; target distance was 9 cm; and chamber vacuum pressure was 660 mm Hg. Bombardment was done twice for each sample. Prothalli were then incubated in darkness for one or two days and GFP-expressing cell prothalli were recovered. GFP-expressing cells, incubated for an additional day in darkness, were used for analysis of chloroplast movement induced by microbeam irradiation. Fluorescence microscopy was done with an Axioplan microscope and a 50 W mercury lamp and a filter set (BP450-490, FT510 and LP515 for an excitation filter, a dichroic mirror and a barrier filter, respectively, Zeiss). These experiments were done five times, and in all cases the results were the same as those described in the text.

Received 22 August; accepted 12 November 2002; doi:10.1038/nature01310.

- Christie, J. M. *et al.* Arabidopsis NPH1: A flavoprotein with the properties of a photoreceptor for phototropism. *Science* **282**, 1698–1701 (1998).
- Kagawa, T. *et al.* Arabidopsis NPL1: A phototropin homolog controlling the chloroplast high-light avoidance response. *Science* **291**, 2138–2141 (2001).
- Jarillo, J. A. *et al.* Phototropin-related NPL1 controls chloroplast relocation induced by blue light. *Nature* **410**, 952–954 (2001).
- Sakai, T. *et al.* Arabidopsis nph1 and npl1: Blue light receptors that mediate both phototropism and chloroplast relocation. *Proc. Natl Acad. Sci. USA* **98**, 6969–6974 (2001).
- Kinoshita, T. *et al.* phot1 and phot2 mediate blue light regulation of stomatal opening. *Nature* **414**, 656–660 (2001).
- Wada, M. & Kadota, A. Photomorphogenesis in lower green plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 169–191 (1989).
- Nozue, K. *et al.* A phytochrome from the fern *Adiantum* with features of the putative photoreceptor NPH1. *Proc. Natl Acad. Sci. USA* **95**, 15826–15830 (1998).
- Kadota, A. & Wada, M. Red light-aphototropic (rap) mutants lack red light-induced chloroplast relocation movement in the fern *Adiantum capillus-veneris*. *Plant Cell Physiol.* **40**, 238–247 (1999).
- Kadota, A., Wada, M. & Furuya, M. Phytochrome-mediated phototropism and different dichroic orientation of Pr and Pfr in protonemata of the fern *Adiantum capillus-veneris* L. *Photochem. Photobiol.* **35**, 533–536 (1982).
- Yatsuhashi, H., Kadota, A. & Wada, M. Blue- and red-light action in photoorientation of chloroplasts in *Adiantum* protonemata. *Planta* **165**, 43–50 (1985).
- Wada, M. & Sci, H. Phytochrome-mediated phototropism in *Adiantum cuneatum* young leaves. *J. Plant Res.* **107**, 181–186 (1994).
- Augustynowicz, J. & Gabrys, H. Chloroplast movements in fern leaves: correlation of movement dynamics and environmental flexibility of the species. *Plant Cell Environ.* **22**, 1239–1248 (1999).
- Nozue, K., Christie, J. M., Kiyosue, T., Briggs, W. R. & Wada, M. Isolation and characterization of a fern phototropin (Accession No. AB037188). A putative blue-light photoreceptor for phototropism (PGR00-039). *Plant Physiol.* **122**, 1457 (2000).
- Christie, J. M., Salomon, M., Nozue, K., Wada, M. & Briggs, W. R. LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): Binding sites for the chromophore flavin mononucleotide. *Proc. Natl Acad. Sci. USA* **96**, 8779–8783 (1999).
- Crosson, S. & Moffat, K. Structure of a flavin-binding plant photoreceptor domain: Insights into light-mediated signal transduction. *Proc. Natl Acad. Sci. USA* **98**, 2995–3000 (2001).
- Kagawa, T. & Wada, M. Phytochrome- and blue-light-absorbing pigment-mediated directional movement of chloroplasts in dark-adapted prothallial cells of fern *Adiantum* as analyzed by microbeam irradiation. *Planta* **198**, 488–493 (1996).
- Smith, H. Phytochrome and light signal perception by plants—an emerging synthesis. *Nature* **407**, 585–591 (2000).
- Bünning, E. & Etzold, H. Über die Wirkung von polarisiertem Licht auf keimende Sporen von Pilzen, Moosen und Farnen. *Ber. Deut. Botan. Ges.* **71**, 304–306 (1958).
- Yatsuhashi, H. & Kobayashi, H. Dual involvement of phytochrome in light-oriented chloroplast movement in *Dryopteris sparsa* protonemata. *J. Photochem. Photobiol. B* **19**, 25–31 (1993).
- Miller, J. H. & Greany, R. H. Determination of rhizoid orientation by light and darkness in germinating spore of *Onoclea sensibilis*. *Am. J. Bot.* **61**, 296–302 (1974).
- Davis, B. D. Bending growth in fern gametophyte protonema. *Plant Cell Physiol.* **16**, 537–541 (1975).
- Tidwell, W. D. & Ash, S. R. A review of selected Triassic to early Cretaceous ferns. *J. Plant Res.* **107**, 417–442 (1994).
- Hasebe, M. *et al.* Fern phylogeny based on *rbcl* nucleotide sequences. *Am. Fern J.* **85**, 134–181 (1995).
- Pryer, K. M., Smith, A. R. & Skog, J. E. Phylogenetic relationships of extant ferns based on evidence from morphology and *rbcl* sequences. *Am. Fern J.* **85**, 205–282 (1995).
- Taylor, T. N. & Taylor, E. L. in *The Biology and Evolution of Fossil Plants* 346–722 (Prentice Hall, Englewood Cliffs, 1993).
- Willis, K. J. & McElwain, J. C. in *The Evolution of Plants* 130–155 (Oxford Univ. Press, Oxford, 2002).
- Yatsuhashi, H. & Wada, M. High-fluence rate responses in the light-oriented chloroplast movement in *Adiantum* protonemata. *Plant Sci.* **68**, 87–94 (1990).
- Kanegae, T. & Wada, M. Isolation and characterization of homologues of plant blue-light photoreceptor (cryptochrome) genes from the fern *Adiantum capillus-veneris*. *Mol. Gen. Genet.* **259**, 345–353 (1998).
- Davis, S. J. & Vierstra, R. D. Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. *Plant Mol. Biol.* **36**, 521–528 (1998).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank W. R. Briggs for critical reading and discussions about the manuscript; H. Nishida for suggestions about the manuscript; M. Hasebe for *H. punctata* genomic DNA; Arabidopsis Biological Resource Center and R. Vierstra for *psmRS-GFP* plasmid; K. Kikuchi for technical advice; and Y. Takahashi, F. Toyama, K. Hara, E. Hayashida, E. Sugiyama and T. Yasuki for their technical assistance. This work was partly supported by Promotion of Basic Research Activities for Innovative Biosciences from Bio-oriented Technology Research Advancement Institution and by Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan to M.W. and by a grant from Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists to H.K., S.C. and Y.S.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.W.
(e-mail: wada-masamitsu@c.metro-u.ac.jp).

Direct activation of RNA polymerase III transcription by c-Myc

Natividad Gomez-Roman*†, Carla Grandori†‡, Robert N. Eisenman‡ & Robert J. White*

* Institute of Biomedical and Life Sciences, Division of Biochemistry and Molecular Biology, University of Glasgow, Glasgow, G12 8QQ, UK

‡ Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, 98109-1024, USA

† These authors contributed equally to this work

The proto-oncogene product c-Myc has a direct role in both metazoan cell growth and division^{1–4}. RNA polymerase III (pol III) is involved in the generation of transfer RNA and 5S ribosomal RNA, and these molecules must be produced in bulk to meet the need for protein synthesis in growing cells⁵. We demonstrate here that c-Myc binds to TFIIB, a pol III-specific general transcription factor, and directly activates pol III transcription. Chromatin immunoprecipitation reveals that endogenous c-Myc is present at tRNA and 5S rRNA genes in cultured mammalian cells. These results suggest that activation of pol III may have a role in the ability of c-Myc to stimulate cell growth.

Cell division requires a critical rate of growth (accumulation of mass), which depends primarily on protein synthesis³. Myc activation drives rapid increases in translation and growth, which precede DNA replication and cell division^{3,6}. Indeed, Myc in some contexts can stimulate protein synthesis and growth independently of cell cycle progression^{7–11}. Conversely, Myc loss diminishes growth and protein synthesis^{3,8,12}. Many candidate Myc target genes encode ribosomal proteins and translation factors^{6,13,14}. Because pol III transcription is tightly linked to growth, we tested whether it responds to c-Myc.

When c-Myc knockout and matched wild-type fibroblasts are compared, the knockout fibroblasts show approximately sevenfold lower expression of pol III transcripts from the B2 repetitive gene family (Fig. 1a); the decrease is specific, as messenger RNA encoding acidic ribosomal phosphoprotein (ARPP) P0 remains constant. Pol III activity is therefore sensitive to the presence of endogenous c-Myc. A tamoxifen-inducible c-Myc-oestrogen receptor (Myc-ER) fusion stimulated B2 expression when introduced into the knockout cells by retroviral transduction¹³. There was an increase of about 1.5-fold over basal levels in the absence of tamoxifen, which reflects leakiness of the conditional protein^{12,15}; however, tamoxifen induced a further increase in pol III transcript levels (Fig. 1b). This response was not observed with the natural variant Myc-S, which lacks the amino-terminal 100 residues of the transactivation domain¹⁵.

Pol III activity was also induced in primary human fibroblasts by Myc-ER expressed from a retroviral vector¹⁵ (Fig. 1c). An intron-specific primer was used in polymerase chain reaction with reverse transcription (RT-PCR) reactions to assay levels of tRNA precursor, which is processed very rapidly and so provides an assay of ongoing transcriptional activity¹⁶. Precursor tRNA^{Leu} levels increase about sevenfold in quiescent cells infected with Myc-ER relative to vector-infected controls; a further increase occurs within 3 h of adding tamoxifen. However, tamoxifen has no effect on cells carrying empty vector. Similar results were obtained with tRNA^{Tyr} (data not shown). Activation of tRNA genes reaches 12-fold in the presence of Myc-ER and tamoxifen, which is a more potent response than the roughly 3.6-fold induction obtained in parallel for the cyclin D2 gene, an established Myc target¹³. These effects are specific, as the level of ARPP P0 mRNA is unchanged. Deletion of residues 106–143 within the c-Myc transactivation domain (Δ Myc-ER) prevents activation of both tRNA and cyclin D2 genes.

Several mechanisms might explain the responsiveness to c-Myc of genes transcribed by pol III. In Rat1 cells, c-Myc was reported to

induce transcription of the BN51 gene, which encodes a subunit of pol III (ref. 17). However, c-Myc did not induce transcription of BN51 in primary human fibroblasts (Fig. 2a). Furthermore, there is no increase in the level of mRNAs encoding subunits of TFIIC2, a pol III-specific factor that is overexpressed in some tumours¹⁶. c-Myc activation can provoke hyperphosphorylation of the retinoblastoma (Rb) protein¹⁸. As Rb represses tRNA synthesis in a phosphorylation-dependent manner^{19,20}, we tested whether the pol III response to c-Myc is Rb-dependent. Transfection of c-Myc expression vector stimulates pol III transcription of the adenovirus VA1 gene in immortalized fibroblasts from either wild-type or Rb-knockout mice, without affecting a co-transfected SV40-CAT control (Fig. 2b). The Rb-related protein p107 also represses pol III transcription²¹; as c-Myc binds and neutralizes p107 (refs 22, 23), we tested whether it continues to activate VA1 in the absence of the entire Rb family (Fig. 2c). Robust and specific VA1 induction is obtained in triple-knockout fibroblasts lacking Rb, p107 and p130. We conclude that pol III regulation by c-Myc does not require the Rb family.

The timing of tRNA induction after activation of Myc-ER matches that of the cyclin D2 gene, a direct target of c-Myc¹³. In contrast, indirect targets generally take longer to respond in this system and S phase entry takes about 17 h (ref. 13). Because of the rapidity of tRNA activation, we considered the possibility of a direct interaction between c-Myc and the pol III machinery. Tamoxifen treatment was carried out in the presence of $2 \mu\text{g ml}^{-1}$ α -amanitin, which specifically blocks pol II transcription: any effect seen under these conditions cannot be a secondary response to induction of

genes transcribed by pol II. Previous studies indicate that Myc-ER protein is active during the time course of the experiment^{13,24}. Whereas mRNA levels of cyclin D2 decline after α -amanitin treatment, the more stable ARPP P0 transcript remains at a steady level throughout the experiment. Although α -amanitin prevented c-Myc activation of the cyclin D2 gene as expected, tRNA^{Leu}, tRNA^{Tyr} and 5S rRNA were induced robustly (Fig. 3). This suggests that c-Myc regulation of pol III templates does not require stimulation of protein-encoding genes, and may therefore be direct.

We used chromatin immunoprecipitation (ChIP) to test whether c-Myc associates with tRNA genes *in vivo*. In serum-starved fibroblasts transduced with empty vector, c-Myc was not detected at the tRNA^{Leu} gene, although TFIIC2 was bound (Fig. 4a). In contrast, c-Myc was found at the tRNA^{Leu} gene in tamoxifen-treated cells carrying the Myc-ER construct. These interactions are specific, as a β -globin gene control showed little binding above the background levels observed with preimmune serum. In cycling fibroblasts, endogenous c-Myc is detected at tRNA^{Leu}, tRNA^{Tyr} and 5S rRNA genes, but not at the ARPP P0 gene (Fig. 4b). These observations are not restricted to fibroblasts, as endogenous c-Myc is also found at 5S rRNA and B2 genes in epithelial cells, but is not detected at the p21^{Waf1} pol II promoter control (Fig. 4c).

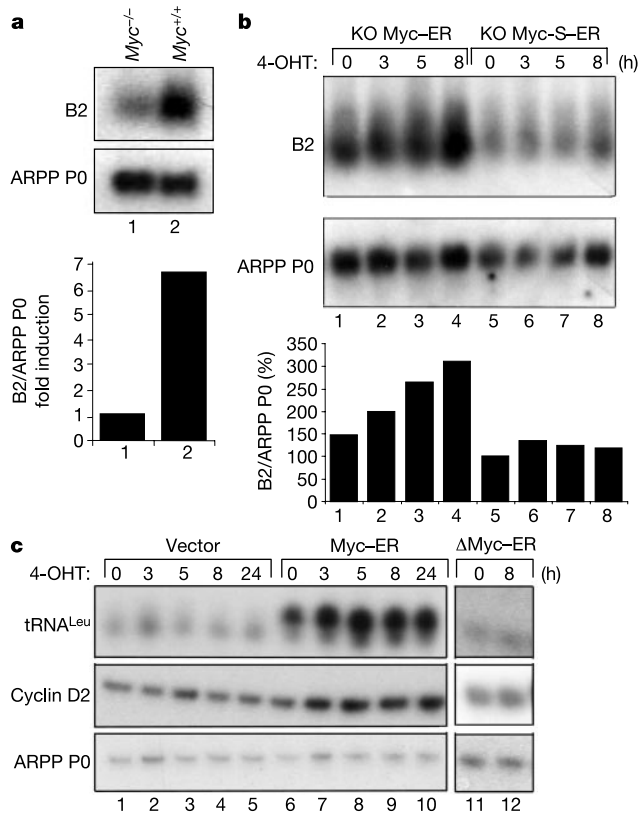


Figure 1 Pol III-transcribed genes respond to c-Myc *in vivo*. **a**, RNA from wild-type or Myc-null fibroblasts¹² was analysed for B2 and ARPP P0 expression by northern blotting. **b**, Myc-null fibroblasts¹² (KO) were transduced with vectors carrying Myc-ER or Myc-S-ER. RNA extracted at the indicated times after treatment with 4-OHT was analysed for B2 and ARPP P0 expression by northern blotting. **c**, WI38 cells were transduced with empty vector or vector carrying Myc-ER or Δ Myc-ER. RNA extracted at the indicated times after 4-OHT treatment was analysed by RT-PCR.

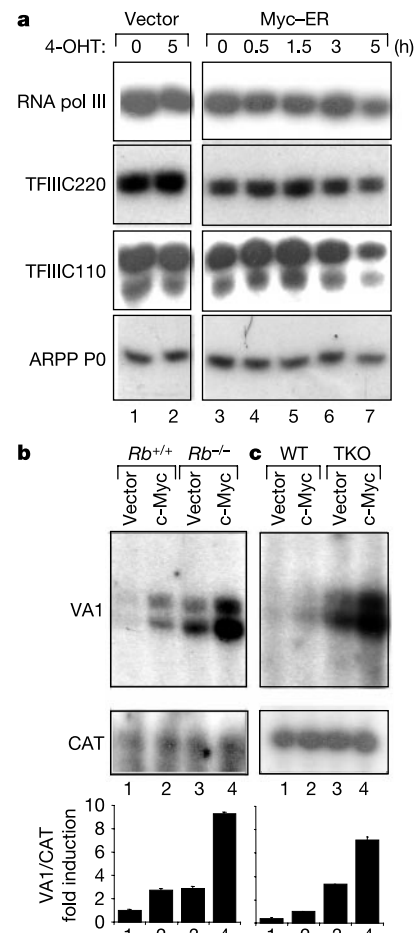


Figure 2 Pol III activation by c-Myc does not require pocket proteins or induction of TFIIC2 or BN51. **a**, RT-PCR of transcripts in WI38 cells transduced with empty vector or Myc-ER, collected at the indicated times after 4-OHT induction. **b**, Primer extension analysis of VA1 and pCAT expression in *Rb*^{+/+} or *Rb*^{-/-} fibroblasts co-transfected with control or c-Myc expression vectors. **c**, As above, except with primary fibroblasts from wild-type (WT) or Rb, p107 and p130 triple knockout (TKO) mice. Bar charts quantify these experiments and independent repetitions, with VA1 expression normalized to the co-transfected pCAT. Error bars indicate standard error of the mean.

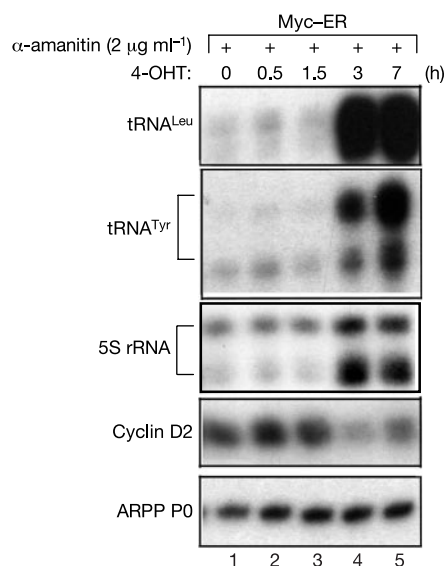


Figure 3 Activation of pol III transcription by c-Myc does not depend on induction of pol II transcription. WI38 cells were transduced with retroviral vector carrying c-Myc-ER, serum-starved and then treated with $2 \mu\text{g ml}^{-1}$ α -amanitin 1 h before induction with 200 nM 4-OHT. RNA was extracted at the indicated times after induction and analysed by RT-PCR for tRNA^{Tyr}, tRNA^{Leu}, 5S rRNA, cyclin D2 and ARPP P0 transcripts.

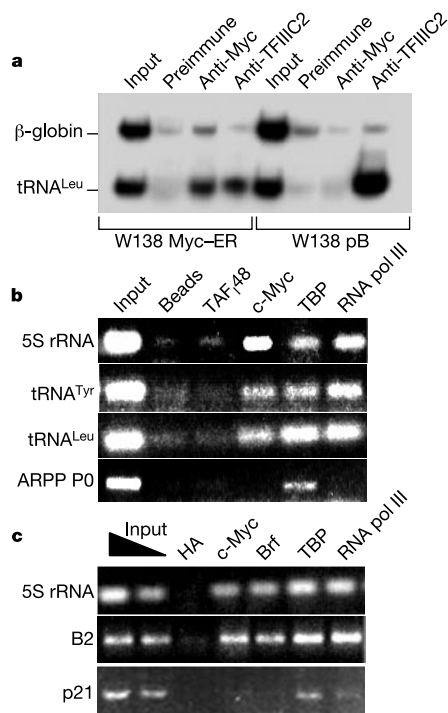


Figure 4 ChIP assays show that c-Myc interacts with tRNA, 5S rRNA and B2 genes *in vivo*. **a**, WI38 cells transduced with Myc-ER or empty vector (pB) were exposed to 4-OHT for 12 h before crosslinking. The ChIP assay used anti-Myc antibody N-262, anti-TFIIB110 antibody 4286 (ref. 16) or preimmune serum. **b**, ChIP assay using Balb/c 3T3 cells and antibodies 9E10 against c-Myc, M-19 against TAF48, BN51 against pol III and MTBP-6 against TBP³⁰. **c**, ChIP assay using ROSE 199 ovarian epithelial cells¹⁶ and antibodies F-7 against HA, 9E10 against c-Myc, 128 against Brf²⁰, BN51 against pol III and MTBP-6 against TBP³⁰.

Specific binding of c-Myc to tRNA genes is also observed in epithelial and B cells (data not shown).

Although ChIP assays show c-Myc located at pol III templates *in vivo*, these genes rarely contain matches to the E-box DNA sequence recognized by Myc. We therefore investigated whether c-Myc regulates pol III transcription by protein-protein interactions in the absence of a DNA recognition site. TFIIB is a plausible target for c-Myc, as it contains TATA box-binding protein (TBP)²⁵, and the transactivation domain of c-Myc is known to bind TBP^{23,26}. Indeed, TBP and the TFIIB-specific subunit Brf1 are retained from extracts by beads carrying glutathione S-transferase (GST) fused to residues 1–262 of c-Myc encompassing its transactivation domain (GST-Myc-N262) (Fig. 5a). Binding is specific, as neither TBP nor Brf1 associate with GST alone or the carboxy-terminal 92 residues of c-Myc containing its DNA-binding and dimerization domain (GST-Myc-C92). Nuclear extracts that have been preincubated with GST-Myc-N262 are depleted of a factor required for pol III transcription (Fig. 5b). Addition of a TFIIB fraction can fully restore tRNA expression, whereas adding pol III does not. We conclude that TFIIB binds specifically to the transcriptional regulatory domain of c-Myc. This is consistent with the fact that induction of tRNA genes *in vivo* requires this domain (Fig. 1).

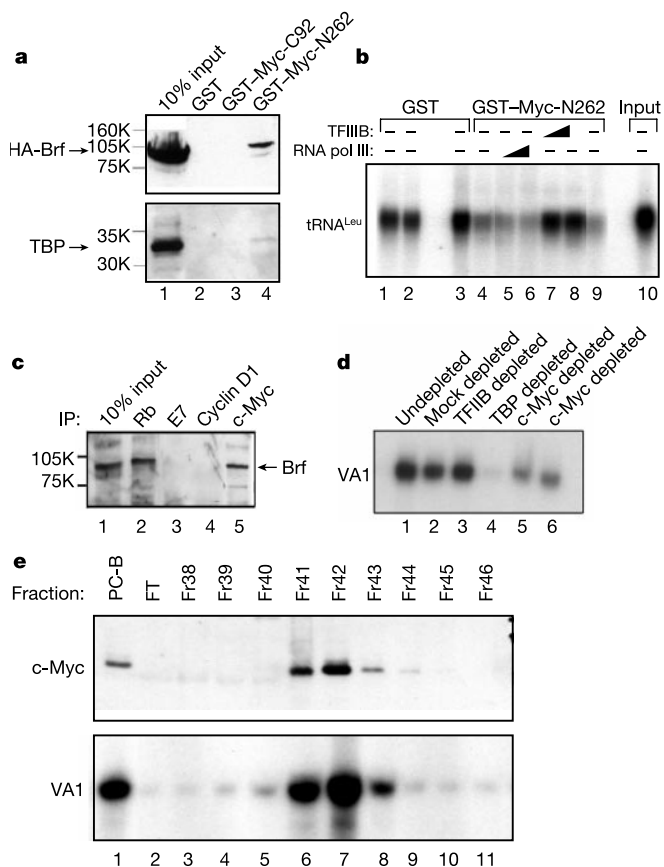


Figure 5 TFIIB interacts with the N-terminal domain of c-Myc. **a**, Pull-down assay with GST, GST-Myc-C92 or GST-Myc-N262, and immunoblotted for HA-tagged Brf1 and TBP. **b**, Transcription of tRNA^{Leu} with extracts depleted of proteins binding to GST or GST-Myc-N262 and supplemented with 2 μl and 4 μl of pol III (lanes 5 and 6) or 2 μl and 4 μl of TFIIB (lanes 7 and 8). **c**, HeLa extract immunoprecipitated using antibodies against Rb, E7, cyclin D1 and c-Myc; precipitates were immunoblotted for Brf1. **d**, HeLa extract mock-depleted or immunodepleted with antibodies against TFIIB, TBP or c-Myc was used to transcribe VA1. **e**, Crude TFIIB was chromatographed on MonoQ, and gradient-eluted fractions were immunoblotted for c-Myc and assayed for TFIIB transcriptional activity with VA1 template.

Co-immunoprecipitation reveals that endogenous c-Myc associates stably with TFIIB at physiological ratios (Fig. 5c). Anti-Rb antibody provides a positive control, as the Rb–TFIIB interaction is well documented^{5,20,21}, whereas antibodies against E7 and cyclin D1 serve as negative controls. As with Rb, the Brf1 subunit of TFIIB immunoprecipitated together with monoclonal antibody against c-Myc. In contrast, anti-Myc antibody did not immunoprecipitate either TFIIC2 or pol III (data not shown). Transcription assays carried out with immunodepleted extract show that binding of endogenous c-Myc to TFIIB is functionally significant. Thus, extracts depleted with anti-Myc antibody display reduced VA1 transcription, albeit less marked than TBP-depleted control (Fig. 5d). In contrast, mock-depletion or immunodepletion of the pol II-specific factor TFIIB has little effect. Independent support for a stable interaction comes from the finding that endogenous c-Myc fractionates together with TFIIB activity on several columns, such as the MonoQ gradient in Fig. 5e. We conclude that the transactivation domain of c-Myc binds stably and specifically to the TFIIB complex.

Identification of direct targets is a current priority of Myc research^{27,28}. The level of tRNA induction by Myc–ER exceeds the weak transactivation often shown by established Myc targets, such as cyclin D2 (Fig. 1c). Furthermore, pol III activity is severely compromised in *c-myc*^{−/−} cells (Fig. 1a), consistent with the interaction between endogenous c-Myc and TFIIB. This interaction may allow recruitment of coactivators and chromatin-modifying activities, or it might block effects of negative regulators of the pol III machinery. The ability of Myc to induce tRNA and 5S rRNA biosynthesis may be expected to impact strongly on cell growth. Previous studies have indicated that a substantial number of putative Myc target genes, transcribed by pol II, are involved in metabolism and translation^{6,13,14}. The capacity to activate both pol II- and pol III-transcribed genes might allow Myc to control growth by coordinately inducing the various components of the protein synthetic apparatus through parallel effects on independent transcription systems. □

Methods

Myc induction

Myc-null rat fibroblasts¹² and WI38 human fibroblasts were transduced as described previously^{13,15} with empty pBABE-puro retroviral vector or vectors carrying Myc–ER, Myc–S–ER, or ΔMyc–ER. They were selected with puromycin for 3 days. Cells were then serum-starved for 48 h to downregulate endogenous *c-myc* expression and then treated with 200 nM 4-hydroxy-tamoxifen (4-OHT).

Gene expression assays

We carried out northern blots as described previously²¹. RT–PCR of tRNA^{Leu}, tRNA^{Tyr}, 5S rRNA, TFIIC220, TFIIC110 and ARPP P0 transcripts was conducted as described previously¹⁶. RT–PCR of cyclin D2 mRNA used primers 5′-CTGCCCCACCTAGATC ATA-3′ and 5′-TCCCTTATGCTGTACTTCAAATAGG-3′; cycling parameters were 94 °C for 30 s, 94 °C for 15 s, 62 °C for 23 s, 72 °C for 30 s, and 72 °C for 5 min. RT–PCR of BN51 mRNA used primers 5′-CCTACACAGTCTGTACCTGG-3′ and 5′-GGCATATTTCGAG TGTCGTTCC-3′; cycling parameters were 95 °C for 150 s, 95 °C for 30 s, 68 °C for 30 s, 72 °C for 30 s, and 72 °C for 5 min. Primer extension analyses of transfected pVA1 and pCAT were carried out as described previously²⁰.

Chromatin immunoprecipitation

Cells (roughly 2×10^7 per immunoprecipitation) were collected and washed with ice-cold PBS and then crosslinked for 10 min at 37 °C in 10 ml 0.5% NP-40/PBS containing 1% formaldehyde. Crosslinked cells were washed with ice-cold 0.5% NP-40/PBS and incubated for 30 min with 40 ml of high salt buffer (0.5% NP-40, PBS, 1 M NaCl). Further washing with 0.5% NP-40/PBS was followed by hypotonic disruption for 30 min in 40 ml of low salt buffer (0.1% NP-40, 10 mM TrisHCl, pH 8.0, 1 mM EDTA, 0.1 M NaCl). After centrifugation at 478g for 10 min, nuclei were obtained with 10 strokes through a 23-gauge needle and re-centrifuged. Nuclei were resuspended in 2.7 ml of low salt buffer and lysed with 300 μl of 20% sarkosyl. Samples were transferred to a sucrose cushion and spun for 10 min at 4,000g. After resuspending the pellet in 3 ml of TE (10 mM Tris, pH 8.0, 1 mM EDTA), the process was repeated. After resuspension in 2 ml of TE, genomic DNA was sheared by sonication (Branson sonifier 250, 10×10 s, duty cycle 30%). Sonicated material was adjusted with 1/10 volume of $\times 11$ NET (1.65 M NaCl, 5.5 mM EDTA, 5.5% NP-40, 550 mM TrisHCl, pH 7.4) and immunoprecipitated overnight at 4 °C with 20 μl (4 μg) of antibody. Protein A or protein G Sepharose beads were added for a further 2 h incubation and then recovered on Polypropylene columns (BioRad). After washing twice with 10 ml of RIPA buffer (50 mM TrisHCl, pH 8.0, 150 mM NaCl, 0.1% SDS, 0.5% deoxycholate, 1% NP-40), twice with 10 ml of LiCl buffer (10 mM TrisHCl, 250 mM LiCl, 0.5% NP-40, 0.5% deoxycholate, 1 mM EDTA, pH 8.0) and twice with 10 ml TE, beads

were transferred to 1.5-ml tubes and immunoprecipitated material was eluted twice with 200 μl of 1% SDS in TE. To isolate precipitated DNA, proteins and antibodies were degraded by proteinase K treatment overnight at 42 °C. DNA was extracted twice using phenol/chloroform/isoamylalcohol and ethanol-precipitated. Immunoprecipitated DNA was quantified by PCR performed on the input and bound fractions, using primers 5′-GAGGACAACGGGGACAGTAA-3′ and 5′-TCCACCAGAAAACTCCAGC-3′ for tRNA^{Leu} and 5′-GACAGGTACGGCTGTCATCA-3′ and 5′-AACGGCAGACTTCTCTCT CAG-3′ for the β-globin control, as described²⁹. PCR of the p21 promoter used primers 5′-GCCTCTTCTGTGCTGCTGA-3′ and 5′-CCAGCCCTTGGATGGTTT-3′; cycling parameters were 95 °C for 3 min, 35 cycles of 95 °C for 1 min, 52 °C for 1 min and 72 °C for 1 min, and 72 °C for 5 min. B2 gene PCR used primers 5′-GGGGCTGGAGAGATGGCT-3′ and 5′-CCATGTGGTGTGCTGGAT-3′; cycling parameters were 95 °C for 3 min, 24 cycles of 95 °C for 30 s, 58 °C for 30 s and 72 °C for 30 s, and 72 °C for 5 min. Other primers and amplification conditions have been described¹⁶.

Pull-down, co-immunoprecipitation and immunodepletion

Pull-down assays were carried out as described previously²¹ with extracts of Chinese hamster ovary (CHO) cells expressing haemagglutinin (HA)-tagged Brf1. These were incubated for 3 h on ice with glutathione-agarose beads carrying GST alone, GST–Myc–C92 or GST–Myc–N262. After pelleting and washing, beads were probed by immunoblotting with antibodies F-7 against HA and 58C9 against TBP. Equal amounts of depleted supernatants were used in transcription assays (Fig. 5b).

Immunoprecipitations were carried out as described²¹ using monoclonal antibodies against Rb (G3–245; Pharmingen), E7 (TVG710Y), cyclin D1 (R-124) and c-Myc (9E10), and precipitates were immunoblotted for Brf1 with antibody 128 (ref. 20). HeLa extract was immunodepleted for 2 h on ice with protein G Sepharose beads alone (mock) or carrying antibodies against TFIIB (C-18), TBP (MTBP-6) or c-Myc (9E10). Equal amounts of immunodepleted extract were then used for transcription (Fig. 5d). Antibodies N-262, 9E10, F-7, TVG710Y, R-124, 58C9, C-18 and M-19 were obtained from Santa Cruz Biotechnologies.

Fractionation and *in vitro* transcription

In vitro transcription assays were carried out as described previously³⁰. Pol III and TFIIB fractions added in Fig. 5b were A25 (1.0) and A25 (0.15), respectively; these fractions, as well as PC-B and PC-C fractions, have been described previously³⁰. Gradient chromatography on MonoQ was conducted as described previously³⁰ with PC-B input; fractions were immunoblotted for c-Myc with 9E10 antibody, and assayed for TFIIB activity in transcriptions reconstituted with PC-C fraction containing pol III and TFIIC.

Received 7 June; accepted 1 November 2002; doi:10.1038/nature01327.

- Dang, C. V. c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol. Cell Biol.* **19**, 1–11 (1999).
- Elend, M. & Eilers, M. Downstream of Myc—to grow or to cycle? *Curr. Biol.* **9**, R936–R938 (1999).
- Schmidt, E. V. The role of *c-myc* in cellular growth control. *Oncogene* **18**, 2988–2996 (1999).
- Grandori, C., Cowley, S. M., James, L. P. & Eisenman, R. N. The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu. Rev. Cell Dev. Biol.* **16**, 653–699 (2000).
- Brown, T. R. P., Scott, P. H., Stein, T., Winter, A. G. & White, R. J. RNA polymerase III transcription: its control by tumour suppressors and its deregulation by transforming agents. *Gene Expr.* **9**, 15–28 (2000).
- Rosenwald, I. B. Upregulated expression of the genes encoding translation initiation factors eIF-4E and eIF-2α in transformed cells. *Cancer Lett.* **102**, 113–123 (1996).
- Iritani, B. M. & Eisenman, R. N. c-Myc enhances protein synthesis and cell size during B lymphocyte development. *Proc. Natl Acad. Sci. USA* **96**, 13180–13185 (1999).
- Johnston, L. A., Prober, D. A., Edgar, B. A., Eisenman, R. N. & Gallant, P. *Drosophila myc* regulates cellular growth during development. *Cell* **98**, 779–790 (1999).
- Schuhmacher, M. *et al.* Control of cell growth by c-Myc in the absence of cell division. *Curr. Biol.* **9**, 1255–1258 (1999).
- Beier, R. *et al.* Induction of cyclin E-cdk2 kinase activity, E2F-dependent transcription and cell growth by Myc are genetically separable events. *EMBO J.* **19**, 5813–5823 (2000).
- Kim, S., Li, Q., Dang, C. V. & Lee, L. A. Induction of ribosomal genes and hepatocyte hypertrophy by adenovirus-mediated expression of c-Myc *in vivo*. *Proc. Natl Acad. Sci. USA* **97**, 11198–11202 (2000).
- Mateyak, M. K., Obaya, A. J., Adachi, S. & Sedivy, J. M. Phenotypes of c-Myc-deficient rat fibroblasts isolated by targeted homologous recombination. *Cell Growth Differ.* **8**, 1039–1048 (1997).
- Coller, H. A. *et al.* Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion. *Proc. Natl Acad. Sci. USA* **97**, 3260–3265 (2000).
- Boon, K. *et al.* N-myc enhances the expression of a large set of genes functioning in ribosome biogenesis and protein synthesis. *EMBO J.* **20**, 1383–1393 (2001).
- Hirst, S. K. & Grandori, C. Differential activity of conditional MYC and its variant MYC-S in human mortal fibroblasts. *Oncogene* **19**, 5189–5197 (2000).
- Winter, A. G. *et al.* RNA polymerase III transcription factor TFIIC2 is overexpressed in ovarian tumors. *Proc. Natl Acad. Sci. USA* **97**, 12619–12624 (2000).
- Greasley, P. J., Bonnard, C. & Amati, B. Myc induces the nucleolin and BN51 genes: possible implications in ribosome biogenesis. *Nucleic Acids Res.* **28**, 446–453 (2000).
- Steiner, P. *et al.* Identification of a Myc-dependent step during the formation of active G₁ cyclin-cdk complexes. *EMBO J.* **14**, 4814–4826 (1995).
- White, R. J., Trouche, D., Martin, K., Jackson, S. P. & Kouzarides, T. Repression of RNA polymerase III transcription by the retinoblastoma protein. *Nature* **382**, 88–90 (1996).
- Scott, P. H. *et al.* Regulation of RNA polymerase III transcription during cell cycle entry. *J. Biol. Chem.* **276**, 1005–1014 (2001).
- Sutcliffe, J. E. *et al.* RNA polymerase III transcription factor IIIB is a target for repression by pocket proteins p107 and p130. *Mol. Cell Biol.* **19**, 4255–4261 (1999).
- Gu, W., Bhatia, K., Magrath, I. T., Dang, C. V. & Dalla-Favera, R. Binding and suppression of the Myc transcriptional activation domain by p107. *Science* **264**, 251–254 (1994).
- Hoang, A. T. *et al.* A link between increased transforming activity of lymphoma-derived MYC mutant

alleles, their defective regulation by p107, and altered phosphorylation of the c-Myc transactivation domain. *Mol. Cell. Biol.* **15**, 4031–4042 (1995).

24. Eilers, M., Schirm, S. & Bishop, J. M. The MYC protein activates transcription of the α -prothymosin gene. *EMBO J.* **10**, 133–141 (1991).

25. Geiduschek, E. P. & Kassavetis, G. A. The RNA polymerase III transcription apparatus. *J. Mol. Biol.* **310**, 1–26 (2001).

26. Hateboer, G. *et al.* TATA-binding protein and the retinoblastoma gene product bind to overlapping epitopes on c-Myc and adenovirus E1A protein. *Proc. Natl Acad. Sci. USA* **90**, 8489–8493 (1993).

27. Cole, M. D. & McMahon, S. B. The Myc oncoprotein: a critical evaluation of transactivation and target gene regulation. *Oncogene* **18**, 2916–2924 (1999).

28. Eisenman, R. N. Deconstructing Myc. *Genes Dev.* **15**, 2023–2030 (2001).

29. Schübeler, D. *et al.* Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human β -globin locus. *Genes Dev.* **14**, 940–950 (2000).

30. White, R. J., Gottlieb, T. M., Downes, C. S. & Jackson, S. P. Mitotic regulation of a TATA-binding-protein-containing complex. *Mol. Cell. Biol.* **15**, 1983–1992 (1995).

Acknowledgements We thank C. Ngouenet and Z. Felton-Edkins for technical assistance, and R. Reeder, S. Cowley and C. Yost for critique of this manuscript. This work was funded by grants from CONACYT-Mexico (N.G.), the National Institutes of Health/National Cancer Institute (C.G. and R.N.E.), and from Cancer Research UK (R.J.W.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.N.E. (e-mail: eisenman@fhcrc.org) or R.J.W. (e-mail: rwhite@udcf.gla.ac.uk).

errata

The role of parasites in sympatric and allopatric host diversification

Angus Buckling & Paul B. Rainey

Nature **420**, 496–499 (2002).

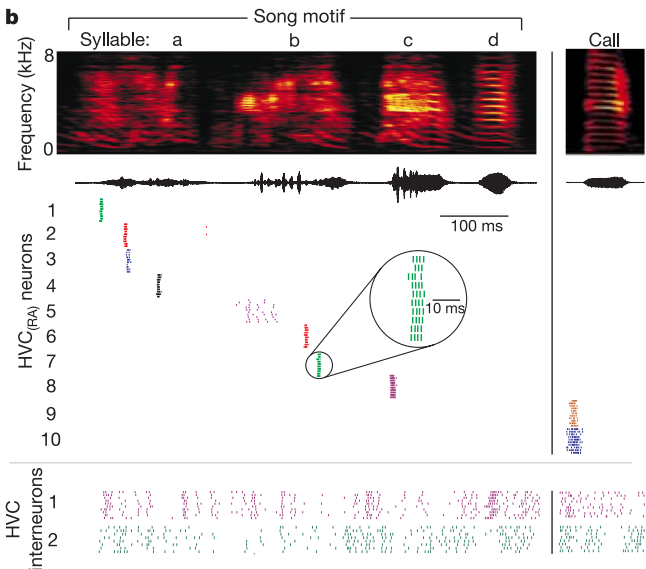
In Fig. 2b of this Letter, the third bar (dark grey) from the left was incorrectly shown. Its allopatric diversity value should have been 0.007 not 0.19. This error does not alter the conclusions of the paper.

An ultra-sparse code underlies the generation of neural sequences in a songbird

Richard H. R. Hahnloser, Alexay A. Kozhevnikov & Michale S. Fee

Nature **419**, 65–70 (2002).

In this Letter, the numbering on the middle panel of Fig. 2b was misaligned. The figure should have appeared as shown.



In the Acknowledgements, F. Nottebohm’s surname was misspelled.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Harajukomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejp.com

naturejobs

Chemistry's clouded view

What a difference a year makes. In March 2001, a survey of the American Chemical Society (ACS) revealed that about 1.5% of its members were unemployed. The results of a similar survey, released late last year, showed a jump in unemployment to 3.3%, the highest figure since the society began tracking employment more than 30 years ago.

The one ray of hope in the survey is that salaries for chemists in their first job after gaining their PhD have climbed by almost 8%. The rest of the analysis is covered in clouds, with more on the horizon. The survey notes that as chemistry employment usually leads economic indicators, rather than follows them, it might be a while before this trend reverses.

Adding further gloom to the scene, the chemical industry increasingly relies on the pharmaceutical sector for employment. But that sector has also been in the doldrums lately, with few successful new products to market, lots of lawsuits and steady consolidation. And now, with the US economy still shaky, many state-funded universities are considering making cuts to the size of their faculty (see *Nature* **421**, 5; 2003).

All of these signs point to longer job hunts. But what can you do to increase your chance of success? The ACS offers some advice. Brush up on your interviewing skills. Consider smaller biotech firms — after ensuring they are solvent. And don't completely rule out pharmaceutical companies; some skills are still in demand, but these tend to be in speciality areas such as computational or synthetic organic chemistry.

Whatever the position, persistence, patience and polishing of both scientific and job-hunting skills should eventually yield results. And, if all else fails, there's always the other p-word — a postdoc.

Paul Smaglik
Naturejobs editor



Contents

SPECIAL REPORT

Europe seeks mobility
for students

p296

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

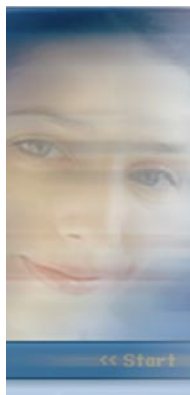
FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Will industry play a bigger role in Europe's aims to promote mobility among postdocs? Susanne Hinck and Quirin Schiermeier investigate.

Working for Unilever was a positive experience for Maria Tabernero. The 25-year-old Spanish biologist spent two years in the laboratory of the chemical and food company in Vlaardingen, the Netherlands, courtesy of the Marie Curie Actions fellowship programme. Marie Curie fellowships are aimed at improving training and mobility for young scientists in Europe — across national borders, but also from academia to industry.

Like many other Marie Curie investigators, Tabernero would now like to continue her career in her home country, preferably at a university. But it remains unclear how these fellowships will help to increase interactions between industry and academia, and industry participation in training young scientists outside their home countries.

Some feel that the inherent premise of fellowships in industry is problematic, as the commercial aspects may run counter to the needs of both sides. Young researchers associate large industrial labs, where most of the non-academic Marie Curie fellowships are offered, with second-class science. And the heads of smaller companies remain closed to the idea of hosting a fellow from another country, out of fear that a short-term fellow could learn and take home more than would be good for the company.

About 85% of the 5,000 young postdocs who were granted an individual Marie Curie fellowship by the European Commission (EC) during the expiring Fifth Framework Programme for research preferred an academic training site. And industrial labs accounted for only 12% of 'host fellowships', which are awarded directly by 500 or so pre-selected host institutions throughout Europe.

Large companies, such as Unilever and Aventis, are keen to promote training opportunities in their labs. Both had stands, for example, at a major European research meeting last October in Brussels. The conference marked the beginning of the Sixth Framework Programme (FP6), from 2002 to 2006, under which the Marie Curie fellowships will be continued.

One goal of FP6 is to improve mobility in the European Research Area. So the overall budget for training and mobility of researchers

has increased by 50% to E1.58 billion (US\$1.64 billion).

But will industry play a bigger role this time? Some researchers say that it certainly has something to offer. "You can expect to work in well-equipped labs, you may get additional marketing and language training, and if you are lucky you will get paid a higher salary than foreseen by the European Commission," says Tabernero.

Nonetheless, industry is unlikely to increase its participation in the Marie Curie programme significantly. In particular, some of the best industrial labs remain unwilling to take part in a scheme that may offer them more risk than benefit.

For example, Peter Stadler, chief executive of one of Germany's most promising biotech start-ups, Artemis in Cologne, doubts that small, highly innovative companies would be prepared to train young guest researchers for fear that core company know-how would wander all over the world.

CATCH-22

Indeed, few small biotech companies have participated in the Marie Curie programme — or even become aware of its existence. They typically say that they lack the resources needed to train and look after fellows, but they are also concerned about confidentiality issues.

"Confidentiality is most important," says Stadler. "We live from selling our intellectual property — who would guarantee that nothing leaks away?"

On the other hand, larger companies, with a high proportion of routine research that is of little relevance in terms of competition, are more likely to value the skills of Marie Curie fellows. Some, indeed, are looking desperately for candidates.

"We've had some excellent fellows in the past 10 years," says Loreto Sheils, a research manager at Unilever's laboratories in Bedford, UK. "But finding good candidates for the industry host fellowships is difficult. One reason may be the reluctance of young scientists to undertake PhD or postdoctoral studies in an industrial environment."

Aventis, which is also offering fellowships, has had similar experiences. Two years ago, for example, the pharmaceutical giant advertised 40 postdoctoral positions in France, 10 of which are still vacant. "We have opportunities mainly for chemists, but these people are much sought after," explains Jutta

Maria Tabernero: industry is well equipped and may even offer postdocs a better salary than academia.





Reinhard-Rupp, who is responsible for scientific affairs at Aventis in Frankfurt, Germany. Filling vacant Marie Curie positions is a particular challenge, she says, because the best potential candidates are put off by the lengthy application procedures when they could probably find better-paid permanent positions elsewhere, without the bureaucracy.

It is particularly difficult for interested small-to-medium-sized enterprises (SMEs) in highly specialized but not very fashionable fields to find fellows. "The few scientists who are looking for industrial training sites are not sufficiently aware about opportunities in SMEs," says Christian Abeln, a Marie Curie supervisor at CAS Software in Karlsruhe, Germany, an SME that develops business software.

CALLING OUT

One attempt to bring interested researchers and SMEs together is the 'Fellow for Industry' initiative, co-funded by the EC and run by the Austrian Bureau for International Research and Technology Cooperation in Vienna. The bureau helps SMEs and researchers with a specific project idea to find appropriate

Peter Stadler fears that commercial sensitivity may hinder postdoctoral studies in industry.

partners in 13 countries, including Israel, Iceland and six future European Union member states in central and eastern Europe.

But despite such initiatives, most Marie Curie fellows prefer academic host institutes as training sites. "It is the freedom to research that counts more than everything else for me," says Joaquin Navarro, a Spanish Marie Curie fellow currently at the Max Planck Institute of Neurobiology in Martinsried, Germany. "I would not really want to work in industry."

Most young scientists are sympathetic. "It is part of the European mentality to think that beauty and purity can solely be found in academic research," says Jonathan

Dando, a member of the advisory board and industry committee of the Marie Curie Fellowship Association (MCFA), an independent group of former fellows who voluntarily provide moral support and practical assistance to newcomers.

"There is a widespread feeling among many young researchers that going to industry would isolate them from the research community, and hence would virtually be scientific suicide," says Dando, a molecular biologist at the Institute Gustave Roussy in Villejuif, France. These fears are not necessarily justified, he believes. "In the life sciences it is not true that industry research cuts you off from the academic community," he says. "Large pharmaceutical companies have a variety of scientific collaborations with small biotech businesses and university groups, so postdocs still have many opportunities to publish."

The MCFA is promoting industry participation in the Marie Curie programme. Last October, for example, it offered all host institutions free space on its website for advertising vacant Marie Curie training opportunities. But the response from industry was poor — only 13 companies made use of the offer. "It may have to do with the difficult economic situation," Dando says. "Many companies are battenning down the hatches."

Why, then, should young scientists apply for a Marie Curie fellowship in industry? One advantage, says Sheils, is that it may help them to find a permanent job if, at some point, they are fed up with the job-hopping typical in academic careers.

But there is no guarantee that an industry fellowship will pave the way to permanent employment. Tired of her futile attempts to find a university position, Tabernero has begun to apply to industrial research departments in Spain. So far, the mention of Marie Curie in her CV has not helped her to find a new job. ■

Susanne Hinck has just finished an internship at Nature's Munich office; Quirin Schiermeier is Nature's German correspondent.



Jonathan Dando believes that postdocs dismiss jobs in industry too readily.

